

AD-A244 725



DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

UNCLASSIFIED		1b. RESTRICTIVE MARKINGS none	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.	
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE JAN 14 1992		5. MONITORING ORGANIZATION REPORT NUMBER(S) A65878-1 16	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) PRO-91-6		7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research	
6a. NAME OF PERFORMING ORGANIZATION Prometheus Inc.	6b. OFFICE SYMBOL (if applicable)	7b. ADDRESS (City, State, and ZIP Code) Building 410 Bolling AFB DC 20332-6448	
6c. ADDRESS (City, State, and ZIP Code) 103 Mansfield Street Sharon, MA 02067	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-90-C-0023		
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Xerox as 7a	8b. OFFICE SYMBOL (if applicable) NM	10. SOURCE OF FUNDING NUMBERS	
8c. ADDRESS (City, State, and ZIP Code) Same as 7b		PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2304
		TASK NO. A9	WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) APPLICATIONS OF APPROXIMATION THEORY IN ANTENNA DESIGN, SIGNAL PROCESSING AND FILTERING (UNCLASSIFIED)			
12. PERSONAL AUTHOR(S) James S. Byrnes			
13a. TYPE OF REPORT final	13b. TIME COVERED FROM 5/1/90 to 9/30/91	14. DATE OF REPORT (Year, Month, Day) 91 November 26	15. PAGE COUNT 59
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) This final report consists of eleven sections. The first is a list of the problems which were considered during the period of performance. Second is the Introduction. Third is the paper <i>Beam-forming applications of polynomials with restricted coefficients</i> , which has appeared in the Proceedings of the 1987 NATO ASI on Electromagnetics (Kluwer Academic Publishers, 1991). Fourth is the paper <i>An ideal omnidirectional transmitting array, and optimal peak factor array, for less than half-wavelength spacing</i> . Fifth is the paper <i>A random variables method for determining the poles of radar targets</i> . Sixth is the paper <i>A computationally efficient notch filter</i> . Seventh is the paper <i>A new rational approximation to digital filters</i> . Eighth is the paper <i>Concerning Prony's method</i> . Ninth is the paper <i>Barker sequences and Littlewood's "two-sided conjectures" on polynomials with ± 1 coefficients</i> . Tenth is the paper <i>A note on rational approximations to the Fresnel integral</i> . Eleventh and last is a list of the Prometheus Inc. personnel who performed the research reported herein, and a list of the papers presented at seminars and conferences. The above-listed papers were authored by James S. Byrnes, W. H. J. Fuchs, W. K. Hayman, T. W. Körner, Donald J. Newman, Gerald Ostheimer, Richard Roy, B. Saffari, and H. S. Shapiro.			
20. DISTRIBUTION / AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS		21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. ARJE NACHMAN		22b. TELEPHONE (Include Area Code) 202/767-4939	22c. OFFICE SYMBOL NM

DD Form 1473, JUN 86

Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

92 1 18 029

**Applications of Approximation Theory in
Antenna Design, Signal Processing and Filtering**

Final Report 26 November 1991

Prometheus Inc.
103 Mansfield Street
Sharon, MA 02067

Contract # F 49620-90-C-0023

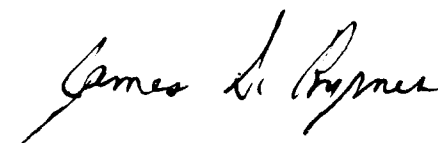
Prepared by:

James S. Byrnes, W. H. J. Fuchs, W. K. Hayman
T. W. Körner, Donald J. Newman, Gerald Ostheimer,
Richard Roy, B. Saffari, H. S. Shapiro

Submitted to:

Dr. Arje Nachman
Mathematics and Information Sciences
Air Force Office of Scientific Research

Submitted by:



James S. Byrnes, President

26 November 1991

Date

92-01111


Applications of Approximation Theory in Antenna Design, Signal Processing and Filtering

Abstract

This final report consists of eleven sections. The first is a list of the problems which were considered during the period of performance. Second is the Introduction. Third is the paper *Beam-forming applications of polynomials with restricted coefficients*, which has appeared in the Proceedings of the 1987 NATO ASI on Electromagnetics (Kluwer Academic Publishers, 1991). Fourth is the paper *An ideal omnidirectional transmitting array, and optimal peak factor array, for less than half-wavelength spacing*. Fifth is the paper *A random variables method for determining the poles of radar targets*. Sixth is the paper *A computationally efficient notch filter*. Seventh is the paper *A new rational approximation to digital filters*. Eighth is the paper *Concerning Prony's method*. Ninth is the paper *Barker sequences and Littlewood's "two-sided conjectures" on polynomials with ± 1 coefficients*. Tenth is the paper *A note on rational approximations to the Fresnel integral*. Eleventh and last is a list of the Prometheus Inc. personnel who performed the research reported herein, and a list of the papers presented at seminars and conferences.



Accession For	
NTIS CRAGI	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail. and/or Special
A-1	

Prometheus Inc.
Final Report
26 November 1991

Section I
Problems considered

A. Rational Approximation Problems, contract #F49620-90-C-0023

Note: problems 1-7 are on the unit circle

1. Approximate a rational function by a lower order rational function, usually in the sup norm. Usually the given rational function has poles only inside $|z| = 1$, and the same is required of the approximating one.
2. Approximate (in sup norm) a given $H(z) = \sum_{n=0}^{\infty} h(n)z^{-n}$ by a rational function

$$R_{MN}(z) = \frac{\sum_{k=0}^M b_k z^{-k}}{\sum_{k=0}^N a_k z^{-k}},$$

where $a_0 = 1$, N and M are given, or $N + M$ is given, or $N - M$ is given, and all poles of R_{MN} are inside $|z| = 1$. Note: Can assume $\sum_{n=0}^{\infty} |h(n)| < \infty$.

3. Same as 2, except $H(z) = \sum_{n=0}^K h(n)z^{-n}$.
- 4-5. Same as 2-3, except only $|H(z)|$ is given and you want $|R_{MN}(z)|$ to approximate $|H(z)|$.
6. Approximate $H(e^{i\omega}) = \omega$ in the L^2 norm by rational functions.
7. Approximate (in sup norm) by rational functions:

$$H(e^{i\omega}) = \begin{cases} 1, & \text{if } 0 \leq \omega \leq b_p - \delta & \text{(Passband)} \\ \text{continuous and linear,} & \text{if } b_p - \delta < \omega < b_p + \delta & \text{(Transition Band)} \\ 0, & \text{if } b_p + \delta \leq \omega \leq \pi & \text{(Stopband)} \end{cases}$$

B. Additional Problems, contract #F49620-90-C-0023

8. $F(\theta) = \sum_{j=1}^P R_j e^{ikd_j \cos \theta}$, k and P are known. You can sample $F(\theta)$ at any θ 's. The R_j 's and d_j 's are real. Devise a sampling procedure so as to determine the R_j 's and d_j 's. This sampling procedure can be adaptive; i.e., take some samples, then compute, then choose further samples based upon these results, etc. Of course, the method should be as (computationally) efficient and stable as possible.
9. Same as 8, except the R_j 's and d_j 's can be complex.
10. Same as 8, except you can only sample $|F(\theta)|$.
- 11-13. Same as 8-10, except you only have a bound on P (say $P \leq 30$).
- 14-19. Same as 8-13, except now the form of F is:

$$F(s) = \sum_{j=1}^P \frac{R_j}{s - s_j} + \sum_{m=-Q}^R C_m s^m, \quad s \text{ complex.}$$

If P, Q, R are not given, 10 is a reasonable bound on each. The poles s_j are usually the most important things to find.

- 20-25. Same as 14-19, except assume s is pure imaginary.
- 26-31. Same as 20-25, except assume $F(s)$ is real.
32. Find a better approximation to the Fresnel integral (see page 8 of the proposal, attached)
33. Find a better solution to the problem of plane wave reflection from an infinite, planar rectangular microstrip-patch array on a grounded dielectric substrate.

Prometheus Inc.
Final Report
26 November 1991

Section II
Introduction

Applications of Approximation Theory in Antenna Design, Signal Processing and Filtering

Introduction

Among the numerous engineering applications of Approximation Theory, the following were those chosen to be the major focus of the research reported herein:

1. Design of *optimal* codes for information transmission and reception,
2. Design of linear functionals (antenna weights) for meeting some pre-defined engineering objective,
3. Digital filter design and implementation.

The most important of these currently, for both military and commercial applications, is that of optimal code design. The wireless communications marketplace is currently congested, a critical problem facing all the service providers in major metropolitan areas today. Though there have been quite a few proposals for increasing the capacity of these systems, one of the currently favored concepts is that of CDMA, or Code-Division Multiple-Access. This spread spectrum technique basically tags each signal in a given frequency channel with a unique code thereby allowing multiple users in the same channel. A critical issue facing designers of such systems is the proper design of these codes for achieving the objectives of improving the quality of service to the customer and the capacity of the overall system. This problem can be cast in a polynomial approximation framework where the objective is to obtain a set of coefficient vectors that have certain orthogonality properties, yet at the same time are it easy to compute with, ie, are ± 1 for example. Advances in the theory behind such sequences would be extremely important in the wireless world of tomorrow, including that of the Air Force and other DoD components. Section IX of this report, regarding Barker Sequences, describes some of our ideas along these lines.

In the design of antenna weights to meet certain engineering objectives, Approximation Theory again comes into play. In many practical systems, arrays of identical antennas are placed in regular lattices. Then a set of weights, one for each antenna, are sought which, when used as linear functionals on the array outputs, perform various tasks which include location of targets/sources, nulling of interferences, minimization of noise power through the system, and many more. In its full generality this is an extremely hard problem. However, for these special regular array structures, the problem can be transformed into a polynomial coefficient design problem. Note that significant hardware simplification can be realized if certain constraints on the coefficients are maintained (eg constant modulus), and this may have a significant impact on cost.

As a note on weight design, in the past one of the main objectives has been the design of polynomials that have unit modulus over only a small region of the unit circle and have minimal modulus elsewhere. The design of such *pencil beams* has been of great interest in phased-array radar technology. With the advent of modern direction finding techniques, the design goals are in the process of being significantly altered. Pencil beams are out, omni-directional sensitivity or gain patterns are in. For such patterns, the objective is to match as closely as possible the output of simple dipole radiators with a large number of elements. The advantage is that there is a significant increase in received energy by employing multiple antennas, energy which is exceedingly important in obtaining accurate direction estimates. Several new Prometheus ideas on antenna weight design are described in sections III, IV, V, and VIII.

Another application of Approximation Theory is in the design of digital filters. Although several of the algorithms developed at Prometheus Inc. require a system of exponentially increasing order for their implementation, and therefore are not immediately amenable to current hardware technology, near-future optical and SAW device technology will provide for FIR filter implementations with many thousands of taps. The approximation technique of Prometheus will be potentially very valuable here. Note that FIR filters are ideal candidates for pipelining, which makes them prime candidates for video bandwidth filters. These will become increasingly important with the advent of HDTV and the next generation wireless networks. The new Prometheus ideas in digital filter design are described in sections VI and VII.

In the Prometheus work on extracting the poles of radar targets by random sampling (section V), a coding problem similar to the Barker sequence problem is being addressed. Here the approach is a stochastic one, but the idea is the same. Obtain a sequence which generates a manifold vector which is maximally orthogonal to all others (in this case with great probability). The reason that astronomical numbers of taps is required is that when dealing with random variables, convergence rates of $O(1/\sqrt{N})$ must be contended with. Again this is simple due to the one-dimensional nature of the solution. Investigations into multidimensional extensions of these concepts could lead to improved convergence rates of the associated algorithms as well as development of new algorithms altogether.

Prometheus Inc.
Final Report
26 November 1991

Section III
Beam-forming applications of polynomials with restricted coefficients

Beam-forming applications of polynomials with restricted coefficients†

James S. Byrnes

Prometheus Inc.

21 Arnold Avenue

Newport, RI 02840

USA

jbyrnes@cs.umb.edu

Also at University of Massachusetts at Boston.

1. Introduction

The basic mathematical question to consider in electronic beam steering with a discrete array consisting of omnidirectional elements spaced at equal increments along a line is how coefficients of a polynomial may be chosen so as to arrive at a desired beam pattern. In numerous applications, these coefficients are required to satisfy certain restrictions, such as a bound on their dynamic range. Here, dynamic range refers to the ratio of the largest to the smallest magnitude. Thus, particularly in null steering, it is often advantageous, or even necessary, for the shading coefficients to all have the same magnitude.

Although the mathematical, statistical, and physical problems that arise in the consideration of array shading have been studied for roughly half a century, many interesting questions remain. In the linear array case under consideration, letting n denote the number of elements, the pattern function $G(z)$ is a polynomial of degree $n - 1$, z is a point on the unit circle, and the shading coefficients are just the n coefficients of this polynomial. An important reason for performing array shading is to shape the pattern function G so that it has low sidelobes and small beamwidth. As is well known, both of these quantities cannot be minimized simultaneously, and

† Research sponsored by the Air Force Office of Scientific Research (AFSC), under Contracts F49620-88-C-0028 and F49620-90-C-0023.

the choice of shading coefficients results in a tradeoff between these two desirable ends.

Electronic beam steering is another fundamental purpose of array shading, and it is this application that we address here. In addition to permitting the rotation of the main response axis of the pattern function, beam steering also allows the simultaneous formation of a number of beams in different directions. In particular, if sources of interference lie at bearings different from that of the desired source, then the signal-to-noise plus interference ratio (SNIR) may be increased dramatically by directing nulls of the pattern function toward these interfering sources, in spite of the fact that the absolute power of the desired signal is thereby reduced. Adaptive techniques have been developed by which array processing systems can electronically respond to an unknown interference environment. However, although the basic adaptive array principles have been known for some time, their application has been limited by hardware constraints and by the lack of sufficiently robust, real-time algorithms. New approaches to this latter consideration are described herein.

There are many cases when constraints must be placed upon the magnitudes of the coefficients of the pattern function. Thus, as explained by Hudson [Hud81], when coefficients are implemented by attenuation, they must be scaled so that the largest modulus is unity, since the amplitude gain for the desired transmission, and even the overall output signal-to-noise ratio (SNR), can be reduced by large coefficients. In discussing main-lobe constraints on optimal arrays, Hudson observes that when a main-lobe null is created, very large shading coefficients are formed, resulting in enhanced output of uncorrelated noise. Hence, size restrictions on the coefficients are again required.

On the other hand, in a situation such as occurs in an adaptive radar receiver after clutter has decayed due to increasing range, so that there will be few and widely spaced target echoes of minimal power compared to a steady jamming source, it is necessary to constrain the adaptive array so that the shading coefficients are prevented from falling to zero. A similar situation occurs in an adaptive antenna using the least mean square (LMS) algorithm, where the shading coefficients will decay to zero if either the signal level falls to zero, or if the reference signal is absent for some reason. One method of controlling this is to substitute the steered gradient system described by Griffiths for the reference signal LMS antenna, but this has the disadvantage of being very sensitive to errors in the assumed direction of the desired signal.

As mentioned earlier, another approach to these questions is to restrict the dynamic range of the shading coefficients. Although an informal rule of thumb for this range appears to be "2 and everyone is happy, 10 and some are happy, 100 and nobody is happy," a formal mathematical study of the relevant properties of polynomials, whose coefficients are thereby restricted, does not seem to have been previously undertaken. An important thrust of the research effort reported herein has been to initiate such a study and to relate to the above applications the large amount of work that has been accomplished by mathematicians on polynomials with restricted coefficients.

Furthermore, there is an intimate relationship between the engineering questions described above and several areas of classical mathematical analysis. Foremost among the problems of mutual interest is the question of how close to constant the modulus of a polynomial can be along some curve, typically the unit circle. This is of great concern to theoretical mathematical analysts because of the fundamental nature of polynomials and the simplicity and intrinsic beauty of the question. It is equally important to engineers working in such fields as array design, adaptive beamforming and null steering, filter design, peak power limited transmitting, and the design of reflection phase gratings. This paper describes our research into both aspects of this remarkable intertwining of the disciplines of pure mathematics and engineering.

2. Mathematical results

Concerning the purely mathematical aspects of our work, note that properties of polynomials with restricted coefficients have been the subject of much fruitful research in twentieth century mathematical analysis. Of particular interest have been polynomials with coefficients ± 1 or complex of modulus one. The study of such functions was apparently initiated by G.H. Hardy (see Zygmund [Zyg59, p. 199]), and furthered by J.E. Littlewood, P. Erdős and others.

For the purposes of this discussion, it will be convenient to introduce the notation of Littlewood [Lit66]. Thus, let F_n and G_n be, respectively, the class of all polynomials of the form

$$f(z) = \sum_{k=0}^n \pm 1 z^k \quad \text{and} \quad g(z) = \sum_{k=0}^n \exp(a_k i) z^k$$

where $|z| = 1$ and the a_k are arbitrary real constants. Clearly, the L^2 norm of g is $\sqrt{n+1}$ for all $g \in G$ (and hence for all $f \in F_n \subset G_n$), and the question "how close can such a g come to satisfying

$$|g| \equiv \sqrt{n+1}?"$$

has long been the object of intense study.

The first qualitative result concerning the above question for G_n was obtained by G.H. Hardy [Zyg59, p. 199], who demonstrated the existence of a positive constant C and a sequence $\{g_n\}$, $g_n \in G_n$, satisfying $|g_n(z)| \leq C\sqrt{n}$ for all n and z . The identical result for F_n was obtained by Shapiro [Sha57] and published by Rudin [Rud59]. Littlewood [Lit62] conjectured that there exist positive constants A and B such that, for any n , there is an $f \in F_n$ ($g \in G_n$) satisfying

$$A\sqrt{n} \leq |f(z)| \leq B\sqrt{n} \quad (A\sqrt{n} \leq |g(z)| \leq B\sqrt{n})$$

for all z , while Erdős conjectured [Erd57] that there is a positive constant C such that for $n \geq 2$, $\|g\|_\infty \geq (1+C)\sqrt{n}$ for all $g \in G_n$ (and hence for all $f \in F_n$). Analogous conjectures for the L^p norms of $g \in G_n$ were settled in a series of papers by Beller and Newman [Bel71, BN71, BN73]. Beller and Newman [BN74] also proved the Littlewood conjecture for polynomials whose coefficients have moduli bounded by 1, after observing that the proof of this result given by Clunie [Clu59] depended on an erroneous result of Littlewood. In [Kör80], Körner was able to modify the result of Byrnes in [Byr77] to prove the Littlewood conjecture for G_n , and then Kahane [Kah80] showed that the Erdős conjecture is false for G_n . These conjectures for F_n remain unresolved.

One approach to the Erdős conjecture for polynomials in F_n is to consider their L^4 norm. For polynomials in G_n , we have the following result:

Theorem 2.1. For each positive integer n , there is a sequence of coefficients $\{c_k\}_{k=0}^n$ such that all $|c_k| = 1$ and

$$\frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{k=0}^n c_k e^{ik\theta} \right|^4 d\theta < (n+1)^2 + 4(n+1)^{3/2}.$$

Proof. We show that, in fact, the Gauss coefficients,

$$c_k = e^{\pi i k^2 / (n+1)},$$

satisfy the required property. Toward that end, note that

$$\sum_{k=0}^n c_k e^{ik\theta} \sum_{m=0}^n \bar{c}_m e^{-im\theta} = n+1 + \sum_{j \neq 0} \left(\sum_{m=0}^n c_m + j \bar{c}_m \right) e^{ij\theta}.$$

Therefore, by Parseval's Theorem, assuming for convenience that n is even,

$$\begin{aligned} \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{k=0}^n c_k e^{ik\theta} \right|^4 d\theta &= (n+1)^2 + \sum_{j \neq 0} \left| \sum_{m=0}^n c_m + j \bar{c}_m \right|^2 \\ &= (n+1)^2 + 2 \sum_{j=1}^n \frac{\sin^2(j^2 \pi / (n+1))}{\sin^2(j \pi / (n+1))} \\ &= (n+1)^2 + 4 \sum_{j=1}^{n/2} \left(\frac{\sin(j^2 \pi / (n+1))}{\sin(j \pi / (n+1))} \right)^2 \\ &\leq (n+1)^2 + 4 \sum_{j=1}^{n/2} \min(j^2, (n+1)^2 / (2j)^2) \\ &\leq (n+1)^2 + 4(n+1)^{3/2}, \end{aligned}$$

where we have used the facts that

$$|\sin jx / \sin x| \leq j, \quad \text{and} \quad |1 / \sin x| \leq \pi / 2x \quad \text{for } 0 < x < \pi / 2.$$

This completes the proof of theorem 2.1.

Another method of constructing polynomials with unimodular coefficients is to form a suitable weighted average of existing ones. For example, we may employ a slight variation of the basic construction in [Byr77] as follows:

For each m , $0 \leq m \leq N^2 - 1$, and for $z = e^{2\pi i \theta}$, let

$$P_m(z) = P_m(\theta) = \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} e^{2\pi i (jk/N + (j+kN)m/N^2)} z^{j+kN}$$

Clearly, each P_m is a polynomial of degree $N^2 - 1$ with coefficients of modulus one. Furthermore, it follows from [Byr77] that, for a suitable small positive ϵ (i.e., of order N^{-2}), $|P_m(\theta)|$ is essentially flat for

$$\epsilon - m/N^2 \leq \theta \leq 1 - \epsilon - m/N^2$$

Now define $P^*(\theta)$ by

$$P^*(\theta) = \sum_{m=0}^{N^2-1} z^{mN^2} P_m(\theta).$$

P^* is a polynomial of degree $N^4 - 1$ with coefficients of modulus one.

Also, by writing

$$P^*(\theta) = \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} e^{2\pi i j k / N} z^{j+kN} \frac{1 - z^{N^4}}{1 - e^{2\pi i (j+kN)/N^2} e^{2\pi i N^2 \theta}}$$

and letting

$$\tilde{\theta} = -N^{-4}(A + BN + CN^2) \quad \text{for } 0 \leq A, B \leq N-1, \quad 0 \leq C \leq N^2-1$$

it is seen that

$$P^*(\tilde{\theta}) = N^2 e^{2\pi i (AB/N - (A+BN)(A+BN+CN^2)/N^4)}$$

so that $|P^*(\tilde{\theta})| = N^2$.

In addition, the essential flatness of $|P^*(\tilde{\theta})|$ in the interval $\epsilon \leq \theta \leq 1 - \epsilon$, where now ϵ is of order N^{-4} , follows as before. However, numerical evidence suggests that $P^*(N^4/2) = O(1)$, a similar situation to that which occurred with the original polynomials [Byr77]. This being the case, P^* is not quite a Kahane-type polynomial, as we had originally hoped.

Note, however, that the above method of constructing $P^*(\theta)$ can also be employed to create new flat spectrum sequences, which are periodic sequences $\{a_k\}_{k=0}^{\infty}$ with the property that their discrete Fourier transform (DFT) has a power spectrum consisting of a very small number (usually one or two) of distinct values. This is because the DFT can be thought of as the values of the polynomial

$$P(z) = \sum_{k=0}^{n-1} a_k z^k,$$

where n is the period, at the n -th roots of unity. Our construction yields polynomials whose spectra are essentially flat at almost all points of the unit circle, not just at the roots of unity. Observe that flat spectrum sequences constructed in this manner satisfy the additional property that all of the terms of the original sequence have the same magnitude. Applications of these concepts to notch filtering and communications are discussed elsewhere in this paper.

Another method of viewing these questions is in the context of interpolation problems. As noted earlier, for any $P \in G_n$, the Parseval Theorem implies that the L^2 norm of P on the unit circle C is $\sqrt{n+1}$. Furthermore, since $|P(z)|^2$ is expressible as $z^{-n} Q(z)$, where Q is of degree $2n$, there can be at most $2n$ distinct points z_k where

$$|P(z_k)| = \sqrt{n+1}$$

Let us call such a set of points an L^2 Interpolating Set for P . A natural question is which, if any, subsets of C consisting of $2n$ points can be an L^2 Interpolating Set for some P of the required form.

In its full generality, this question appears to be quite difficult. For $n = 1$, it is trivial to show that $S = \{a, b\}$ is an L^2 Interpolating Set if and only if $b = -a$. For arbitrary n , observe that for $S = \{z_k\}_{k=1}^{2n}$ to be an L^2 Interpolating Set, the coefficients of P must be chosen so that

$$Q(z) - (n+1)z^n = \alpha \prod_{k=1}^{2n} (z - z_k),$$

where α is a constant of modulus one. Furthermore, the coefficient of z^n on the left side of this equation vanishes, so the same must be true on the right side. Clearly, this will be a very rare occurrence, so that most sets will not be L^2 Interpolating Sets. In fact, it is not at all obvious that for $n > 1$, there exist *any* L^2 Interpolating Sets. Thus far, we are only able to show that if S is to be such a set, its elements cannot be too close to each other. More precisely,

Theorem 2.2. For any n , there is an $\epsilon > 0$ such that no S of the form

$$S = \{e^{i\theta_k}\}_{k=1}^{2n}, \quad \text{with } |\theta_k| \leq \epsilon \text{ for } 1 \leq k \leq 2n,$$

is an L^2 Interpolating Set for any $P \in G_n$.

Proof. Assume the contrary. Fix n . Then, for any $\epsilon > 0$, there is a set

$$S = S(\epsilon, n) = \{e^{i\theta_k}\}_{k=1}^{2n}, \quad \text{with } |\theta_k| \leq \epsilon \text{ for } 1 \leq k \leq 2n,$$

such that S is an L^2 Interpolating Set for some P , say

$$P_{\epsilon, n}(z) = \sum_{k=0}^n a_{\epsilon, k} z^k, \quad \text{with all } |a_{\epsilon, k}| = 1.$$

Choose a sequence of position ϵ 's, say $\{\epsilon_j\}_{j=1}^{\infty}$, approaching 0.

For each k , $0 \leq k \leq n$, the sequence $\{a_{\epsilon_j, k}\}_{j=1}^{\infty}$ is bounded, and all terms of each of these $n+1$ sequences have modulus one. By the standard method of choosing a convergent subsequence for one k at a time, we can find a strictly increasing sequence of positive integers

$$\{m_j\}_{j=1}^{\infty}, \quad \text{and a set } \{a_k\}_{k=0}^n$$

of complex numbers all of modulus one, such that

$$\{a_{\epsilon_{m_j}, k}\}_{j=1}^{\infty}$$

converges to a_k for every k , $0 \leq k \leq n$. Since

$$|P_{\epsilon_{m_j}}(e^{i\theta})| - \sqrt{n+1}$$

can't change sign for

$$|\epsilon_{m_j}| \leq \theta \leq \pi,$$

we can assume, by taking another subsequence if necessary, that either

$$|P_{\epsilon_{m_j}}(e^{i\theta})| - \sqrt{n+1}$$

is always positive or always negative for

$$|\epsilon_{m_j}| < \theta \leq \pi.$$

Suppose the former (the argument being the same in the latter case), and define

$$P_0(z) = \sum_{k=0}^n a_k z^k.$$

Clearly,

$$\{P_{\epsilon_{m_j}}(z)\}_{j=1}^{\infty}$$

converges uniformly to $P_0(z)$ on $|z| = 1$, so that

$$|P_0(e^{i\theta})| \geq \sqrt{n+1} \quad \text{for } 0 < \theta \leq 2\pi.$$

Since the L^2 norm of P_0 is $\sqrt{n+1}$, this is impossible, and the proof of theorem 2.2 is complete.

Also of interest is the locations of the zeroes of polynomials with unimodular coefficients. This is directly related to many other problems

discussed herein and has obvious importance in the choice of pattern functions for null steering. To quantify this question, let $r_j e^{i\alpha_j}$, $1 \leq j \leq n$, be the zeros of $P_n \in G_n$, normalize P_n so that the coefficient of z^n is 1, and define

$$\lambda_n = \max_j \min_j |1 - r_j| \quad \text{and} \quad \lambda_{n,q} = \max \left(\sum_{j=1}^n |1 - r_j|^q \right)^{1/q},$$

where the maximum is taken over all such $P_n(z)$.

Since any $P_1(z) = z - e^{i\alpha_1}$ for some real α_1 , it is obvious that

$$\lambda_1 = \lambda_{1,q} = 0.$$

Considering the case $n = 2$,

$$P_2(z) = z^2 - (r_1 e^{i\alpha_1} + r_2 e^{i\alpha_2})z + r_1 r_2 e^{i(\alpha_1 + \alpha_2)}$$

so that

$$r_1 r_2 = |r_1 e^{i\alpha_1} + r_2 e^{i\alpha_2}| = 1.$$

Assume that $r_1 \geq 1$. Since

$$1 = |r_1 e^{i\alpha_1} + r_2 e^{i\alpha_2}| \geq r_1 - r_2 = r_1 - 1/r_1,$$

the maximum value for $r_1 - 1/r_1$ (hence the maximum value for $r_1 - 1$) is achieved when $r_1 - 1/r_1 = 1$, or

$$r_1 = \frac{1 + \sqrt{5}}{2} \quad \text{and} \quad r_2 = \frac{2}{1 + \sqrt{5}}.$$

In this case,

$$r_1 - 1 = \frac{\sqrt{5} - 1}{2} \quad \text{and} \quad 1 - r_2 = \frac{\sqrt{5} - 1}{1 + \sqrt{5}} < \frac{\sqrt{5} - 1}{2},$$

so that

$$\lambda_2 = \frac{\sqrt{5} - 1}{\sqrt{5} + 1} = \frac{3 - \sqrt{5}}{2}.$$

Also,

$$\lambda_{2,2}^2 = \max_r ((r - 1)^2 + (1 - 1/r)^2).$$

By an elementary calculus argument, it is seen that this maximum occurs for $r = (\sqrt{5} + 1)/2$. Thus,

$$\lambda_{2,2} = \sqrt{5 - 2\sqrt{5}}.$$

We leave as an open question the behavior of other values of λ_n and $\lambda_{n,q}$.

3. Applications

As mentioned in section 1, applications of polynomials with restricted coefficients abound in the engineering world. Those which we focus on herein include null steering, adaptive beamforming, notch filtering, peak power limited transmitting, and the synthesis of low peak-factor signals and flat spectrum sequences.

Several new designs of analytic null steering algorithms for linear arrays are described in [BN88]. Two of them, the β -Technique and the Positive Coefficient Model, allow for placing an arbitrary number of nulls in arbitrary directions while maintaining main beam and sidelobe level control. A method of incorporating these deterministic null steering techniques into existing adaptive algorithms is proposed. The resulting Direct Adaptive Nulling System offers the possibility of significant increases in array performance at very little cost.

A major reason for combining deterministic methods with existing techniques is that arrays must ordinarily deal with significant random noise. In these cases, one has no a priori information about the direction or nature of such unwanted signals. Thus, in such applications, as well as in cases where advance knowledge of jammer characteristics is lacking, indirect statistical methods are unavoidable, although their efficiency may be greatly increased by combining them with analytic approaches.

There exist applications, however, where much is known in advance about the characteristics of both the desired signals and the undesired noise. This is especially true where one has control of the generation of these waveforms. Thus, in the case where one system is producing both offensive signals (i.e., searching for and homing in on targets) and defensive signals (i.e., identifying and tracking incoming weapons), so that mutual interference becomes a predominant concern, the problem is almost exclusively deterministic in nature. In such cases, robust and computationally efficient analytic algorithms controlling both the individual performance of the offensive and defensive signals and the interactive jamming between them are crucial to mission success.

A related problem is the determination of optimal shading coefficients for a conformal array. As is well known, using various measures of optimality, this is a computational problem of order n^3 , where n is the number of array elements. Thus, the computational load will be reduced by a factor of 8 if the coefficients may be restricted to be real. Circumstances where this occurs are described in [Byr88a]. A different method of improving computational efficiency, namely a convex programming approach, will be an important focus of further research.

Another interesting application of our concepts is to notch filters. A nearly ideal notch filter employing coefficients of equal magnitude is given in [Byr88b]. The construction is based upon earlier work of the author involving polynomials with restricted coefficients [Byr77]. The fundamental idea employed in [Byr88b] to construct a notch filter with a single notch may be combined with the concept of an n -nomial [Byr73] to produce nearly ideal filters with multiple notches. Furthermore, as noted elsewhere, zero coefficients do not affect the dynamic range, so that these multi-notch filters maintain the property of having unit dynamic range.

In addition to their use in the construction of notch filters, Byrnes Polynomials [Byr77, Kah80, Kör80] have potential applications to the design of peak power limited transmitters and the synthesis of low peak-factor signals and flat spectrum sequences. In transmitter design, for example, one is often faced with a peak power constraint. Under various conditions, the transmitter output may be modeled as a polynomial. Here the maximum modulus of the polynomial on the unit circle represents the peak power, while the L^2 norm of the polynomial is the average power. Thus, the classical engineering problem of minimizing the peak-to-average ratio becomes the mathematical question of minimizing the ratio of the sup norm to the L^2 norm of a polynomial on the unit circle.

In the trivial case where one frequency is to be transmitted (i.e., the polynomial can be a monomial), clearly the ideal value 1 for the peak-to-average ratio is achieved, and the polynomial is indeed of constant modulus on the unit circle. For the more interesting and practical case of transmitting many linearly increasing frequencies, it is usually desired to transmit each frequency at the same power, which should be as large as possible. As the power of each individual frequency is represented by the modulus of the corresponding coefficient, the mathematical question naturally arises of how close to constant the modulus of a polynomial with equimodular coefficients can be on the unit circle.

More precisely, if n pure tones are transmitted with frequencies of

the form $f_0 + k\Delta$, where f_0 is the fundamental frequency and Δ is the increment, then the waveform is

$$\begin{aligned} x(t) &= \sum_{k=0}^{n-1} A_k \cos(2\pi(f_0 + k\Delta)t + \theta_k) \\ &= S(t) \cos(\arg S(t) + 2\pi f_0 t). \end{aligned}$$

Here,

$$S(t) = \sum_{k=0}^{n-1} A_k e^{i\theta_k} e^{i2\pi k \Delta t},$$

θ_k = phase, and A_k = power in k th tone.

As mentioned, almost always all frequencies are transmitted with equal power, so that $A_k \equiv 1$. To minimize the peak power of $x(t)$, the maximum (over t) of $|x(t)|$ must be minimized (over θ_k). It is relatively straightforward to see that the exact problem is to obtain

$$\min_{\theta_k} \max_t \left| \sum_{k=0}^{n-1} e^{i\theta_k} e^{i2\pi k \Delta t} \right|,$$

a job which is performed by the Byrnes polynomials [Byr77] in nearly ideal fashion.

The adaptation of such polynomials to these problems is important, since in applications like the Link 11 Communications System, the average power is usually maintained at one tenth or less of its theoretical ideal to prevent transmitter overload. Employing concepts such as those described above should yield a significant reduction in the peak-to-average ratio, thereby allowing a large increase in average power, hence a more efficient communications system. These considerations also show that the Byrnes construction has direct application to the synthesis of low peak-factor signals.

Now consider the problem of designing a flat spectrum sequence $\{a_k\}_{k=0}^{\infty}$ as defined earlier. These sequences have direct use in such diverse areas as concert hall acoustics, the quieting of an object's response to radar and active sonar, and speech synthesis. Schroeder [Sch85] presents many of the fascinating details of these applications.

As we observed, the DFT can be thought of as the values of the polynomial

$$P(z) = \sum_{k=0}^{n-1} a_k z^k,$$

where n is the period, at the n -th roots of unity. The Byrnes construction [Byr77] yields polynomials whose spectra are essentially flat at almost all points of the unit circle, not just at the roots of unity. Furthermore, they have the additional property that all of the terms of the original sequence, $\{a_k\}$, have the same magnitude. Applications of these concepts to notch filtering and communications are discussed elsewhere.

In our final application, we have begun to exploit the great success of J.P. Kahane [Kah80] in solving the Littlewood conjecture. As we note in section 2, Kahane showed that there indeed exist polynomials with unimodular coefficients whose modulus is essentially constant on the unit circle. It is our opinion that the breakthrough of Kahane was due to his ingenious use of randomness and probability in his construction. Behind his and previous approaches was the idea of Gauss, viz. the "Gauss Sums." To put it quite simply, we feel that Littlewood's problem was vanquished by the "equation"

$$\text{Kahane} = \text{Gauss Sums} + \text{Probabilistic Choices}$$

Our idea is to exploit the Kahane breakthrough by developing methods to judiciously make the "Probabilistic Choices" referred to above, and thereby convert Kahane's "randomized" proof into a constructive one. This would not only result in exciting new mathematics, but would also be directly applicable to several important engineering problems. In addition to the areas of peak power limited transmitting and flat spectrum sequences discussed earlier, such polynomials would find immediate use in the design of reflection phase gratings, and therefore would be employable in solving concert hall acoustics problems and in quieting the response of an object to sonar or radar. Another potential application of this "educated randomness" construction is in the synthesis of multi-element omnidirectional beam patterns.

In the concert hall acoustics application of reflection phase gratings, it is desired to design the ceiling so that sound is widely scattered *except* in the specular direction. As described earlier, in the context of notch filter design, the Byrnes polynomials [Byr77] place a null in any given direction while the coefficients maintain their other desirable properties of being both flat spectrum and low correlation sequences. Thus, they might even be preferable to the Kahane polynomials in this context. This also appears to apply to monostatic radar, where the null would be placed in the direction of the radar. For bistatic radar, on the other hand, the receiver direction is often unknown. Thus, if a construction based upon the Kahane polynomials could be employed, radar energy would be reflected

equally in all directions, thereby reducing the probability that there would be enough energy reflected in any particular direction to enable detection. A possible undersea application of these ideas occurs in the design of baffles used to quiet machinery noise from submarines, in an attempt to prevent the noise from escaping the hull. Note that our constructions would complement the coatings that are already in use or are being designed to attack these problems, since these coatings provide uniform attenuation. Furthermore, surface structures based upon the Byrnes polynomials would have the highly diffusing property over a large set of frequencies. It is not yet clear whether the Kahane polynomials also yield this important property. The design of two-dimensional arrays so that energy may be scattered with equal intensity over the solid angle is also of considerable interest. It appears that a straightforward product formulation gives the desired results for the Byrnes polynomials, but the situation is not so clear for the Kahane polynomials. We continue to focus our research on the many fascinating questions raised in this final paragraph.

4. Bibliography

- [Bel71] E. Beller. Polynomial extremal problems in L^p . *Proc. Amer. Math. Soc.*, 30:249–259, 1971.
- [BN71] E. Beller and D.J. Newman. An l_1 extremal problem for polynomials. *Proc. Amer. Math. Soc.*, 29:474–481, 1971.
- [BN73] E. Beller and D.J. Newman. An extremal problem for the geometric mean of polynomials. *Proc. Amer. Math. Soc.*, 39:313–317, 1973.
- [BN74] E. Beller and D.J. Newman. The minimum modulus of polynomials. *Proc. Amer. Math. Soc.*, 45:463–465, 1974.
- [BN88] J.S. Byrnes and D.J. Newman. Null steering employing polynomials with restricted coefficients. *IEEE Trans. Antennas and Propagation*, 36(2):301–303, 1988.
- [Byr73] J.S. Byrnes. L^2 approximation with trigonometric n -nomials. *J. of Approx. Th.*, 9:373–379, 1973.
- [Byr77] J.S. Byrnes. On polynomials with coefficients of modulus one. *Bull. London Math. Soc.*, 9:171–176, 1977.

- [Byr88a] J.S. Byrnes. The minimax optimization of an antenna array employing restricted coefficients. *Scientia*, 1:25–28, 1988.
- [Byr88b] J.S. Byrnes. A notch filter employing coefficients of equal magnitude. *IEEE Trans. Acoustics, Speech, Signal Processing*, 36(11):1783–1784, 1988.
- [Clu59] J.C. Clunie. The minimum modulus of a polynomial on the unit circle. *Quarterly J. of Math.*, 10:95–98, 1959.
- [Erd57] Paul Erdős. Some unsolved problems. *Mich. Math. J.*, 4:291–300, 1957.
- [Hud81] J.E. Hudson. *Adaptive Array Principles*. Peter Peregrinus Ltd., New York, 1981.
- [Kah80] Jean-Pierre Kahane. Sur les polynomes a coefficients unimodulaires. *Bull. London Math. Soc.*, 12:321–342, 1980.
- [Kör80] T.W. Körner. On a polynomial of Byrnes. *Bull. London Math Soc.*, 12:219–224, 1980.
- [Lit62] J.E. Littlewood. On the mean values of certain trigonometric polynomials, II. *Illinois J. Math*, 6:1–39, 1962.
- [Lit66] J.E. Littlewood. On polynomials $\sum \pm 1z^m, \sum e^{a_m i} z^m$, $z = e^{i\theta}$. *J. Lon. Math. Soc.*, 41:367–376, 1966.
- [Rud59] W. Rudin. Some theorems on Fourier coefficients. *Proc. Amer. Math. Soc*, 10:855–859, 1959.
- [Sch85] M.R. Schroeder. *Number Theory in Science and Communication*. Springer-Verlag, Berlin, second edition, 1985.
- [Sha57] H.S. Shapiro. Extremal problems for polynomials and power series. Sc.M. thesis, Massachusetts Institute of Technology, 1957.
- [Zyg59] A. Zygmund. *Trigonometric Series*. Cambridge University Press, second edition, 1959.

Prometheus Inc.
Final Report
26 November 1991

Section IV

An ideal omnidirectional transmitting array, and optimal peak factor array, for less than half-wavelength spacing

An ideal omnidirectional transmitting array, and optimal peak factor array, for less than half-wavelength spacing

James S. Byrnes †
Prometheus Inc.
21 Arnold Avenue
Newport, RI 02840

Abstract

We give an explicit construction of an ideal omnidirectional transmitting array, and optimal peak factor array, for the case of a linear array of identical omnidirectional elements with uniform spacing of less than half-wavelength. The construction is based upon the Byrnes polynomials, introduced by the author in 1977.

Introduction

A consideration which is often important in the synthesis of an antenna pattern is the *peak factor*, which is the ratio of the peak to average power of the array. For a line array of equally spaced omnidirectional elements, the mathematical model for the pattern is a polynomial whose coefficients are the *weights*, or *shading coefficients*, of the array. The quantity to be synthesized in this case is the magnitude of $P(z)$ on the unit circle $\mathcal{C} = \{z : |z| = 1\}$, and the peak factor is now the ratio of the *sup* norm to the L^2 norm of P .

Clearly the peak factor is bounded below by 1, and the classical problem is to make it as close to 1 as possible. In addition to arising in peak power limited transmitting and other aspects of antenna design, the identical question occurs in digital filtering. Furthermore, the same polynomial problem has been the object of intense study by mathematical analysts for more than fifty years, including such notable mathematicians as P. Erdős [3], G.H. Hardy, J.P. Kahane [4], T. Körner [5], J.E. Littlewood [6], and D.J. Newman [7].

A parallel concern, to engineers and mathematicians alike, is the question of synthesizing various antenna patterns (i.e., constructing polynomials with specified moduli on the unit circle) when certain restrictions are placed upon the coefficients. When transmitting, for example, one usually wants to maximize the total power output of the array, which is achieved when each individual element broadcasts at full power. Thus, in this case, the coefficients must have the same magnitude. In addition, the usual peak power limitation means that the output power must be as close to constant as possible in all directions, i.e., $|P(z)|$ should be as close to constant as possible on \mathcal{C} . The fundamental result in this specific area is that of Kahane [4]:

Theorem. There is an absolute constant $c > 0$ such that for each positive integer n , there is a polynomial $P(z)$ of degree n with unimodular coefficients which satisfies

$$|P(z)| = \sqrt{n} + E \quad \text{for all } z \in \mathcal{C}, \quad \text{where } |E| \leq cn^{\frac{3}{10}} \log n. \quad (1)$$

Since any such P has $n+1$ coefficients of modulus 1, so that its L^2 norm is obviously $\sqrt{n+1}$, this shows that the ideal peak factor of 1 is indeed achievable asymptotically, even when the coefficients are required to have the same magnitude. Note that Kahane's theorem shows that the ratio of the maximum modulus of P to the minimum modulus of P can be asymptotically 1 as well, a stronger result than the peak factor one just mentioned. As an aside, we mention that the corresponding problem when the coefficients are required to be ± 1 is one of the important unsolved problems in this area of mathematical analysis.

Although Kahane's beautiful result was greeted with much enthusiasm by the mathematical community, it has not found application in the engineering problems discussed above. This is because his proof is an *existence* one: he shows probabilistically that such polynomials must occur. Thus far such functions have not been actually constructed. This too is an important outstanding question, in both mathematics

† The author is also Professor of Mathematics at the University of Massachusetts at Boston

Research sponsored by the Air Force Office of Scientific Research (AFSC), under Contracts F49620-88-C-0028 and F49620-90-C-0023

and engineering. The purpose of this note is to give an *explicit* construction of an ideal omnidirectional transmitting array, and optimal peak factor array, for the case of a linear array of identical omnidirectional elements with uniform spacing of less than half-wavelength. In certain circumstances such close spacing can cause a notable increase in mutual coupling. This physical problem will not be addressed here.

Results

As is well known, for a linear array of n equally spaced identical omnidirectional elements the array factor is the polynomial

$$P(z) = \sum_{j=0}^{n-1} a_j z^j, \quad z = e^{2\pi i u}, \quad u = \frac{d \sin \theta}{\lambda} \quad (2)$$

Here, the a_j are the shading coefficients, d is the array spacing, λ is the wavelength, and θ is the angle of incidence ($\theta = 0$ is broadside).

Note that for $d = \lambda/2$, generally considered the ideal case, $2\pi u$ goes from $-\pi$ to π as $\sin \theta$ goes from -1 to 1 , so that z traverses \mathcal{C} completely. However, if $d < \lambda/2$, then a portion of \mathcal{C} will be omitted by z . This is precisely the property that we now exploit.

In fact, the construction follows immediately from earlier work of the author [2]. The crucial property of the polynomials introduced there, and later called *Byrnes polynomials* [1, 5], is given by part (ii) of the basic theorem in [2]:

Theorem. For n a positive integer, let

$$P(z) = \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} \exp(2\pi i j k n^{-1}) z^{j+kn}, \quad z = e^{2\pi i u}. \quad (3)$$

Then for any ε , $n^{-1} < \varepsilon < \frac{1}{2}$, we have $|P(z)| = n + E$ for all u , $\varepsilon < |u| \leq \frac{1}{2}$, where $|E| < 1 + 2\pi^{-1} + 5(\pi\varepsilon)^{-1}$.

To clarify the significance of this result in the present context, suppose for simplicity that $d = \lambda/4$, so that z is now traversing exactly $1/2$ of \mathcal{C} as $\sin \theta$ goes from -1 to 1 . The theorem immediately yields the correct choice of coefficients, gotten by merely changing z to $-z$ in (3). Hence, the optimal polynomial is

$$P(z) = \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} (-1)^{j+kn} \exp(2\pi i j k n^{-1}) z^{j+kn}, \quad z = e^{2\pi i u}, \quad u = \frac{\sin \theta}{4}. \quad (4)$$

This gives $|P(z)| = n + E$, $|E| < 1 + 22\pi^{-1} < 9$ for all θ . Note that the degree of P is $n^2 - 1$, so that its L^2 norm is n and it represents an array with n^2 elements. Furthermore, the error E is *uniformly* bounded independent of n , so that we indeed have an optimal peak factor array. In addition, the coefficients of P are unimodular, so that it also represents an ideal omnidirectional transmitting array. These properties are illustrated in Figures 1.a-1.e, showing polar plots of $|P(z)|$ in Equation (4) for various choices of n and for $-\pi \leq \theta \leq \pi$. Observe that for small n the actual array patterns exhibit an error that is considerably smaller than that obtained for the general case considered in [2]. Finally, the spread (i.e., the difference between the maximum and minimum, expressed in dB's) of $|P(z)|$ as a function of the number n^2 of elements is shown in Figure 2. This graph clearly illustrates the desired flatness of $|P(z)|$, as described above.

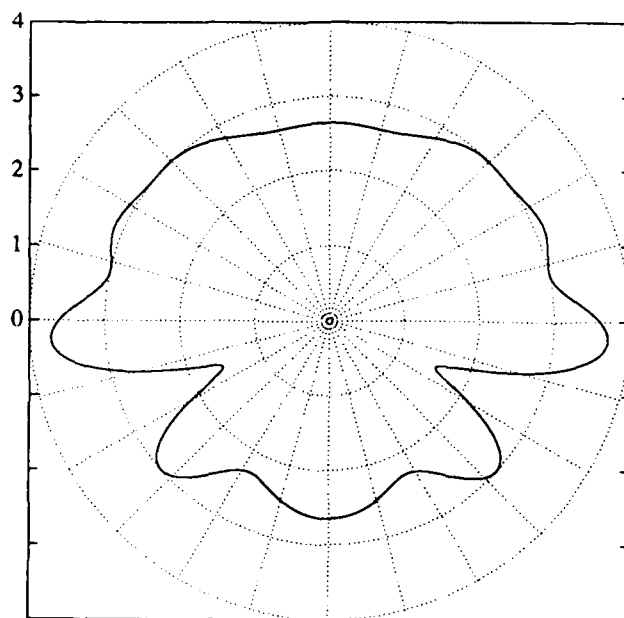


Figure 1.a: 9 elements ($n=3$)

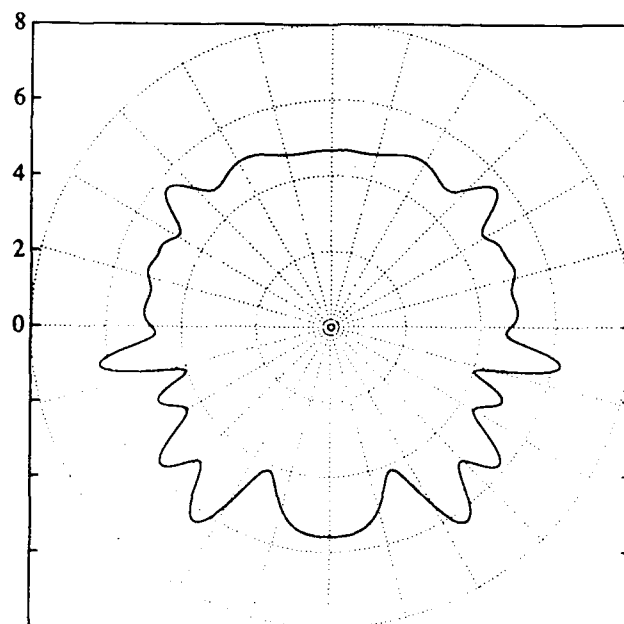


Figure 1.b: 25 elements ($n=5$)

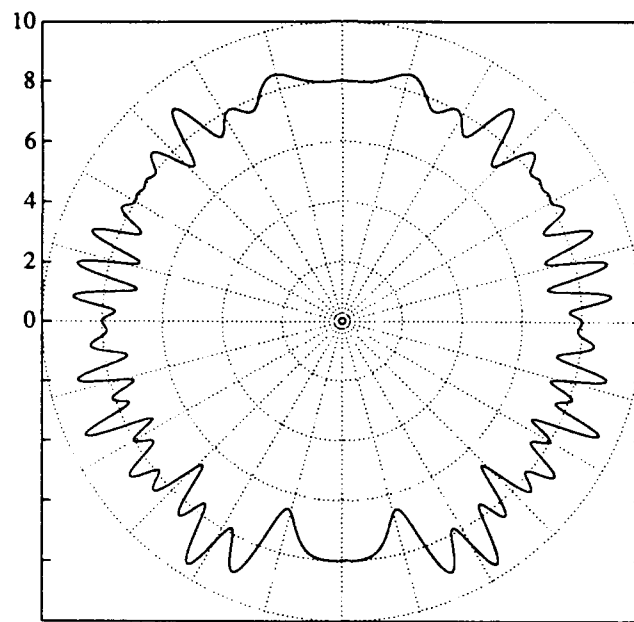


Figure 1.c: 64 elements (n=8)

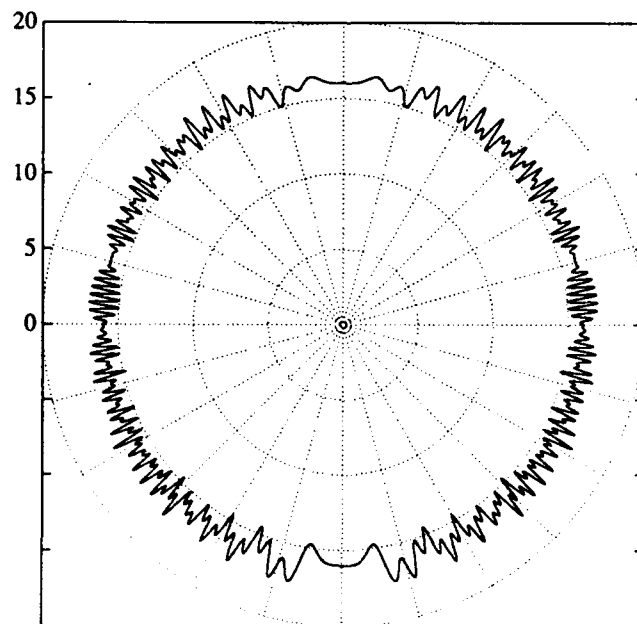


Figure 1.d: 256 elements (n=16)

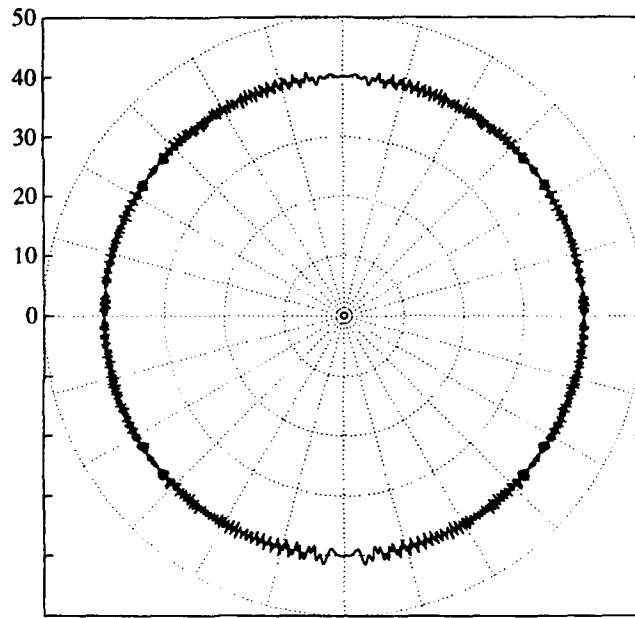


Figure 1.c: 1600 elements (n=40)

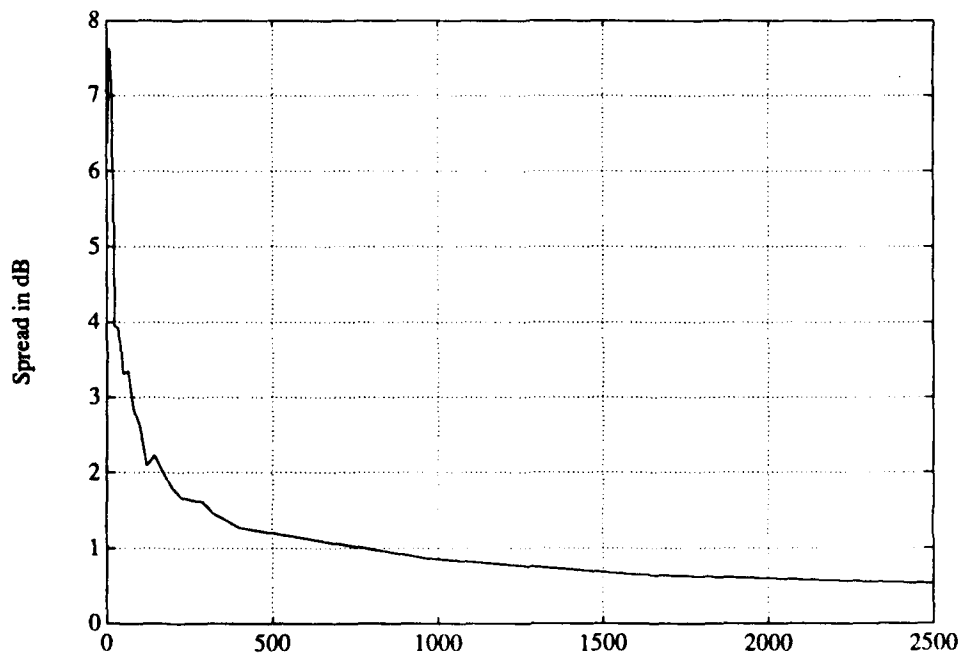


Figure 2: Spread of $|P(z)|$ as a function of the number of elements.

Bibliography

- [1] G. Benke, "On the maximum modulus for a certain class of unimodular trigonometric polynomials," in *Recent advances in Fourier analysis and its applications*, pp. 73-90, J. S. Byrnes and J. L. Byrnes, eds., Kluwer, 1990.
- [2] J. S. Byrnes, "On polynomials with coefficients of modulus one," *Bull. London Math. Soc.*, vol. 9, pp. 171-176, 1977.
- [3] P. Erdős, "Some unsolved problems," *Mich. Math. J.*, vol. 4, pp. 291-300, 1957.
- [4] J.-P. Kahane, "Sur les polynômes à coefficients unimodulaires," *Bull. London Math. Soc.*, vol. 12, pp. 321-342, 1980.
- [5] T. W. Körner, "On a polynomial of Byrnes," *Bull. London Math. Soc.*, vol. 12, pp. 219-224, 1980.
- [6] J. E. Littlewood, "On the mean values of certain trigonometric polynomials," *J. London Math. Soc.*, vol. 36, pp. 307-334, 1961.
- [7] D. J. Newman, "An L^1 extremal problem for polynomials," *Proc. Amer. Math. Soc.*, vol. 16, pp. 1287-1290, 1965.

Prometheus Inc.
Final Report
26 November 1991

Section V
A random variables method for determining the poles of radar targets

A Random Variables Method for
Determining
the Poles of Radar Targets*

T.W.Körner
Prometheus Inc.
21 Arnold Avenue
Newport, RI 02840
Also at Trinity Hall, Cambridge.

G. Ostheimer
Prometheus Inc.
Also at St. Andrews University.

J.S. Byrnes
Prometheus Inc.
Also at University of Massachusetts at Boston.

April 26, 1991

Abstract

Hidden periodicities may be more easily detected if sampling takes place at random times.

*Research supported by AFOSR, under contract F49620-90-C-0023

1 Introduction

When the transient scattered field of a radar target is approximated by an exponential sum

$$y(t) = \sum_{i=1}^M b_i e^{s_i t}, 0 \leq t \leq t_y, \quad (\text{notation from [HS89]})$$

the s_i are the dominant resonant poles of the target. As they represent the resonant frequencies of the target, they are ordinarily the most important information available in the transient field. Since these frequencies are aspect independent they give a signature of the object, and thus can be used as a means of target identification [Per75].

The study of efficient methods for estimating the poles of radar targets, apparently initiated by Kennaugh and Moffatt [KM65], remains an active area of research. The techniques employed thus far include Prony's method [BM75], the pencil-of-function approach [S⁺80, JSW83], and the E-pulse method [RCN87, HS89]. Each of these methods have their particular advantages and drawbacks, with none being close to ideal in any global sense. For example, one problem with Prony's method is its reliance on uniformly sampled sinusoidal functions. In the E-pulse method, long integration times, which could reduce the noise component in the smoothed signal $x(t)$, may have the adverse effect of increasing the sensitivity of the estimate poles to the noise in $x(t)$. Thus, new mathematical methods of estimating the error in the approximation of the transient scattered field of a radar target by an exponential sum, and of recovering the resonant frequencies of the target from this sum, are called for.

In this paper we consider the following special case.

Problem 1 Suppose $f : \mathbb{R} \rightarrow \mathbb{C}$ is a function of the form

$$f(x) = \sum_{m=1}^M a_m \exp(i\lambda_m x) \quad (1)$$

with $a_m \in \mathbb{C}$ and $\lambda_m \in \mathbb{R}$ [$1 \leq m \leq M$], but that we do not know the values of the a_m and λ_m or even of M . The result of making an observation at a point $x_n \in \mathbb{R}$ is given by

$$\tilde{f}(x_n) = f(x_n) + e_n \quad (2)$$

where e_n is some unknown random error. We are allowed to make observations at x_1, x_2, \dots, x_N and wish to estimate the a_m and λ_m from the resulting observations $\tilde{f}(x_1), \tilde{f}(x_2), \dots, \tilde{f}(x_N)$.

As it stands the question is not fully defined. We make the following remarks.

(1) Unless the error e_n is zero it is unreasonable to expect to recover all the λ_m however small the associated a_m . Even if there is no error we would expect a method which claimed to recover λ_m for all non-zero a_m to be numerically unstable.

(2) It is unreasonable to expect to recover those λ_m outside a previously specified set Λ composed of a finite number of intervals. Dirichlet's Theorem tells us that, given any function f of the form (1) together with $x_1, x_2, \dots, x_N \in \mathbb{R}$ and any $\epsilon > 0$ and $K > 0$ we can find $\lambda'_m \in \mathbb{R}$ such that

$$|\lambda'_m - \lambda_m| > K$$

but

$$\left| \sum_{m=1}^M a_m \exp(i\lambda'_m x_n) - \sum_{m=1}^M a_m \exp(i\lambda_m x_n) \right| < \epsilon$$

for all $1 \leq n \leq N$.

(3) Next we observe that if f is given by

$$f(t) = \exp(i\lambda t) - \exp(i(\lambda + \eta)t)$$

then $|f(t)| < \epsilon$ for all $|t| < |\eta|^{-1}\epsilon/2$. It is thus unreasonable to seek a method which does not require previous knowledge of

$$\lambda^* = \min_{r \neq s} |\lambda_r - \lambda_s|.$$

(4) We also need to know something about the errors e_n . In what follows we shall assume that the e_n are independent identically distributed random variables with mean $E[e_n] = 0$ and variance $E[e_n^2] = \sigma^2$. We shall further assume that the distribution of e_n is Gaussian (since we work in \mathbb{C} this means that $\arg e_n$ is uniformly distributed on $[0, 2\pi)$ and $\operatorname{Re} e_n$ is a real Gaussian random variable) but our arguments apply, with hardly any change, whenever the distribution of e_n is reasonably well behaved. More complicated models

of errors exist in which the e_n are not independent. I believe that the method proposed will produce similar results for some of these as well.

(5) We must have some idea, not only of the cost of single computation, but of the cost of making a single observation. The relative cost may lead us to prefer a method requiring many computations but few observations or vice versa. The absolute cost may make some methods impractical.

So far the considerations we have raised apply to all methods. Our discussion will make two specific assumptions which will not always be satisfied in practice.

(6) We assume that the choice of x_1, x_2, \dots, x_N precedes any of the observations. In other words our method is not adaptive (though, of course, it could be incorporated into an adaptive scheme).

(7) Finally we assume that the x_1, x_2, \dots, x_N may be chosen freely from the reals.

In order to illustrate the remarks above and to provide a comparison with the methods we shall propose, consider the following standard procedure. Suppose that f is as in equation (1) and we know that the λ_m all lie well within $\Lambda = [-\alpha, \alpha]$. Choose $T > 0$ and set

$$x_n = -T + 2nT/N.$$

We now obtain the observations \tilde{f} given by equation (2) and compute the 'Fourier transform'

$$\tilde{F}(-\alpha + 2r\alpha/N) = N^{-1} \sum_{n=1}^N \tilde{f}(x_n) \exp(-i(-\alpha + 2r\alpha/N)x_n)$$

for $r = 1, 2, \dots, N$, using the fast Fourier transform. If N is reasonably large (specifically, that $\sqrt{N}/\log N$ is large compared to $\sigma/\min_{1 \leq m \leq M} |a_m|$), it is clear that $\tilde{F}(-\alpha + 2r\alpha/N)$ will not differ appreciably from

$$F(-\alpha + 2r\alpha/N) = N^{-1} \sum_{n=1}^N f(x_n) \exp(-i(-\alpha + 2r\alpha/N)x_n).$$

In particular, provided that λ^*T is large, the graph of F will exhibit M typical regions of disturbance centered on the λ_m of maximum amplitude $|a_m|$ and typical width on the order T^{-1} . Thus, provided we take N roughly of the same size as $T\alpha$, we can locate the λ_m to a precision of about T^{-1} . This

method thus requires on the order of $\alpha\delta^{-1}$ observations and $\alpha\delta^{-1}\log(\alpha\delta^{-1})$ computations to locate the frequencies λ_m to within δ .

In the next section we give an informal description of our proposed method along the same lines. Section 3 compares our method with the one just described. The following section contains several examples employing our method.

2 Description

Choose $T > 0$ and a reasonably large integer N (for simple problems $N = 400$ might be sufficient). Choose x_1, x_2, \dots, x_N *at random* in the interval $[-T, T]$. More formally, set $x_n = X_n$ where X_1, X_2, \dots, X_N are independent, identically distributed random variables each uniformly distributed on $[-T, T]$. (Of course, the interval $[-T, T]$ may be replaced by any other interval of length $2T$.)

Now suppose that f is as in equation (1) and we have the observations \tilde{f} given by equation (2). We define an 'approximate Fourier transform' $\tilde{F}(\lambda)$ for a frequency λ by

$$\tilde{F}(\lambda) = N^{-1} \sum_{n=1}^N \tilde{f}(X_n) \exp(-i\lambda X_n).$$

Observe that (keeping λ fixed) $\tilde{F}(\lambda)$ is a random variable given by

$$\tilde{F}(\lambda) = N^{-1} \sum_{n=1}^N Y_n$$

where the

$$Y_n = (f(X_n) + e_n) \exp(-i\lambda X_n)$$

are themselves independent, identically distributed random variables taking values in \mathbb{C} . We note that

$$EY_n = E(f(X_n) \exp(-i\lambda X_n)) = F(\lambda)$$

where

$$F(\lambda) = \frac{1}{2T} \int_{-T}^T f(t) \exp(-i\lambda t) dt$$

and that

$$\begin{aligned} E|Y_n|^2 &= E(|f(X_n)|^2 + e_n^* f(X_n) + e_n f(X_n)^* + |e_n|^2) \\ &= E|f(X_n)|^2 + E|e_n|^2. \end{aligned}$$

If $\lambda^* T$ is sufficiently large the cross terms in the evaluation of $E|e_n|^2$ will be negligible, and so setting $\tau^2 = \text{var}(Y_n)$, we will have

$$\tau^2 \leq \sum_{m=1}^M |a_m|^2 + \sigma^2 + \eta \quad (3)$$

where η is negligible. Applying the central limit theorem we see that

$$N^{1/2} \tilde{F}(\lambda) = N^{-1/2} \sum_{n=1}^N Y_n$$

is approximately Gaussian with mean $F(\lambda)$ and variance τ^2 . In other words

$$\tilde{F}(\lambda) = F(\lambda) + E(\lambda) \quad (4)$$

where $E(\lambda)$ is approximately equal to the normal random variable Z with mean zero and variance $\tau^2 N^{-1}$. We think of $E(\lambda)$ as noise which is partly natural, coming from the errors e_n , and partly artificially induced by the random choice of the sample points. In the same way we interpret the inequation (3) by saying that the variance τ^2 has a natural component σ^2 and an artificially induced component $\tau^2 - \sigma^2$. Because of the nature of the Gaussian distribution we would not be surprised to find $|Z|$ of size about $\tau N^{-1/2}$ or even about $2\tau N^{-1/2}$ but it would be extremely surprising to find $|Z|$ of size $5\tau N^{-1/2}$ or greater.

With this in mind let us fix a $K \geq 3$. The probability that $|Z| \geq K\tau N^{-1/2}$ is $1 - p(K)$ where $p(K)$ is given by

$$\frac{2}{\sqrt{2\pi}} \int_K^\infty \exp(-x^2/2) dx.$$

Since

$$\begin{aligned} F(\lambda) &= \frac{1}{2T} \int_{-T}^T f(t) \exp(-i\lambda t) dt \\ &= \sum_{m=1}^M a_m \frac{\sin((\lambda_m - \lambda)T)}{(\lambda_m - \lambda)T} \end{aligned} \quad (5)$$

we see that, if $|\lambda_m - \lambda|$ is reasonably large compared with T^{-1} for all m , then $\tilde{F}(\lambda)$ will be smaller than $(K+1)\tau N^{-1/2}$ with probability at least $1 - p(K)$. On the other hand, provided that $\lambda^* T$ is reasonably large, if $|a_r| \geq 3K\tau N^{-1/2}$ then for λ close to λ_r the term

$$\frac{\sin((\lambda_r - \lambda)T)}{(\lambda_r - \lambda)T}$$

will dominate in the expansion

$$\tilde{F}(\lambda) = \sum_{m=1}^M a_m \frac{\sin((\lambda_m - \lambda)T)}{(\lambda_m - \lambda)T} + E(\lambda)$$

with probability at least $1 - p(K)$.

So far we have looked at $\tilde{F}(\lambda)$ for a single value of λ . Now suppose that we are given Λ , the union of a finite number of disjoint intervals of total length $|\Lambda|$. Provided R is reasonably large compared with the number of intervals making up Λ , we can find R points $\nu_1, \nu_2, \dots, \nu_R \in \Lambda$ such that if $\lambda \in \Lambda$ then $|\lambda - \nu_r| < |\Lambda|/R$ for some $1 \leq r \leq R$. It is, of course, not true that the 'errors' $E(\nu_r)$ are mutually independent but the simplest possible estimate shows that

$$|E(\nu_r)| \leq K\tau N^{-1/2} \quad \text{for all } 1 \leq r \leq R \quad (6)$$

with probability at least $1 - Rp(K)$. If (6) holds then the same kind of considerations as applied in the previous paragraph and in the introduction will apply, provided we have $|\Lambda|R^{-1}$ of size about $(6T)^{-1}$ or smaller. If $|a_m| \geq 3\tau N^{-1/2}$ and λ_m is well within Λ we shall see a typical region of disturbance centered on λ_m of amplitude roughly $|a_m| \pm K\tau N^{-1/2}$ and width of the order T^{-1} standing out from the surrounding noise. We may or may not detect regions associated with smaller $|a_m|$ but, provided we ignore all ν_r with $|\tilde{F}(\nu_r)| \leq (K+1)\tau N^{-1/2}$ we shall not obtain any 'false positives'.

A well-known estimate gives

$$\begin{aligned} & \frac{2}{\sqrt{2\pi}} \int_K^\infty \exp(-x^2/2) dx \\ & \leq \frac{2}{K\sqrt{2\pi}} \int_K^\infty x \exp(-x^2/2) dx \\ & = \frac{2}{K\sqrt{2\pi}} \exp(-K^2/2). \end{aligned}$$

Thus, the procedure of the last paragraph will locate all the λ_m in the search region for which $|a_m| \geq a^*$ to a precision of about δ (and not produce false positives) with a probability of failure less than ϵ provided the following relations hold. (Here $u \succeq v$ is to be read as $u \geq Cv$ for some numerical constant C .)

$$\begin{aligned}\delta &\succeq T^{-1} \\ T^{-1} &\succeq |\Lambda|R^{-1} \\ a^* &\succeq K\tau N^{-1/2} \\ \epsilon &\succeq RK^{-1}\exp(-K^2/2).\end{aligned}$$

Thus we need

$$\epsilon N^{1/2} a^* \tau^{-1} \exp(N(a^* \tau^{-1})^2/2) \succeq R \succeq |\Lambda|\delta \quad (7)$$

for our method to perform as desired. The inequality(7) tells us how we must vary N , the number of sample points. Since the sample points are not in an arithmetic progression we cannot use the fast Fourier transform and our method will require on the order of RN computations.

3 Discussion

The reader who is worried by the probability ϵ of error should observe that, if we fix the other parameters, then N increases slower than $\log(\epsilon^{-1})$. Thus an essentially trivial increase in the number of sample points and computations will reduce the probability of failure inherent in the method below e.g. the probability of some serious undetected computer error. It should also be noted that if $\sigma > 0$ then *any* method must inevitably have a strictly positive probability of error.

Let us now compare our method with that discussed at the end of the introduction. Let us fix ϵ and a^* ; for example, we might take

$$a^* = \min_{1 \leq m \leq M} (a_m).$$

In the introduction we took $\Lambda = [-\alpha, \alpha]$ and saw that the standard method required on the order of $|\Lambda|\delta^{-1}$ sample points and $|\Lambda|\delta^{-1} \log(|\Lambda|\delta^{-1})$ computations to locate the frequencies λ_m to within δ . Our method, which

allows Λ to be the finite union of intervals, requires, at most, on the order of $(\log(|\Lambda|\delta^{-1}))^{1/2}$ sample points and on the order of $|\Lambda|\delta^{-1}(\log(|\Lambda|\delta^{-1}))^{1/2}$ computations. We note, for later use, that the choice of Λ can be made *after* the observations.

There are various remarks we should make at this stage.

(1) Although we have achieved a substantial reduction in the number of sample points required, the number of computations has been only slightly reduced. In future research, we expect to show that this weakness can, to a large extent, be overcome.

(2) If we keep a_1, a_2, \dots, a_M fixed but allow the λ_m to vary, it is not hard to see that the accuracy δ and the power of discrimination λ^* vary in step with each other. Thus our method requires at most, on the order of $(\log(|\Lambda|\lambda^*))^{1/2}$ sample points and on the order of $|\Lambda|\lambda^*(\log(|\Lambda|\lambda^*))^{1/2}$ computations. Although we have chosen the random variables X_n to have uniform probability distribution on $[-T, T]$, smoothing them might well produce better discrimination in practice. For example, if we take the X_n to be independent, identically distributed random variables each with density function

$$\begin{aligned} g(x) &= T^{-1}|T - x| \text{ for all } |T - x| \leq T \\ &= 0 \text{ otherwise} \end{aligned}$$

and proceed as before, our method and the supporting argument are essentially unchanged except that equation 5 becomes

$$F(\lambda) = \sum_{m=1}^M a_m \frac{(\sin((\lambda_m - \lambda)T/2))^2}{((\lambda_m - \lambda)T/2)^2}.$$

We have thus localized the disturbance associated with the frequency rather better than before.

(3) The 'detection/noise' ratio

$$\rho^* = \frac{a^*}{N^{1/2}(\sum_{m=1}^M |a_m|^2 + \sigma^2)^{1/2}}$$

only decreases as fast as $N^{-1/2}$. Thus to reduce ρ^* by a factor of L while keeping everything else unchanged requires us to multiply the number of sample points, and so, also, the number of computations by a factor of order L^2 .

At first sight this does not seem very good, but a simple example shows that the rates of growth in the last sentence cannot be improved. Suppose that we simplify our problem so that equation (1) becomes, simply,

$$f(x) = a$$

with $a \in \mathbb{C}$. Then equation (2) gives

$$\tilde{f}(x_n) = a + e_n$$

and our problem reduces to finding whether $a = 0$ or not. Simple statistical considerations show that, in order to be reasonably confident of detecting $a \neq 0$ when $|a| \geq a^*$ and not declaring $a \neq 0$ when, in fact, $a = 0$, we need a 'detection/noise' ratio

$$\rho^{**} = \frac{a^*}{N^{1/2}\sigma}$$

which is larger than 1. Since we must read the values of $\tilde{f}(x_n)$ in order to use them we must make at least N computations. Just as before, to reduce ρ^{**} by a factor of L while keeping everything else unchanged requires us to multiply the number of sample points, and so, also, the number of computations by a factor of order L^2 . It is, of course, true that ρ^{**} only involves the 'natural noise level' σ while ρ^* involves a noise level with an 'artificial component' of order $(\sum_{m=1}^M |a_m|^2)^{1/2}$. In future work we shall sketch a way of getting round this.

(4) We turn now to the estimation of a_m . The obvious way of doing this is to guess $a_m = \tilde{F}(\nu_j)$ where ν_j is the point at which $|\tilde{F}(\nu_k)|$ is largest within the disturbance associated with λ_m . Other schemes are possible, but the reader should remember that the random errors $E(\nu_k)$ of equation (4) are not independent. With our simple scheme there are two sources of error. The first source is the distance of ν_j from λ_m . This error will be of order $|a_m|\delta^2$ and, provided δ has already been chosen quite small, will not be important. (In any case, once we have located λ_m as lying near ν_k we can always calculate $\tilde{F}(\nu)$ for a group of closely spaced points near ν_k without adding noticeably to the computational load or to the number of sample points required.)

We are thus only worried by the second source of error, the noise $E(\lambda)$. Almost exactly the same considerations as applied to the problem of detection in note (3) show that this error will be on the order of $N^{-1/2}\tau$ for our method and that any method what so ever must have errors in the estimation of a_m

of order at least $N^{-1/2}\sigma$. The last sentence of (3) thus applies with ρ^* and ρ^{**} replaced by

$$\epsilon^* = \frac{\text{error level}}{N^{1/2}\tau} \text{ and } \epsilon^{**} = \frac{\text{error level}}{N^{1/2}\sigma}.$$

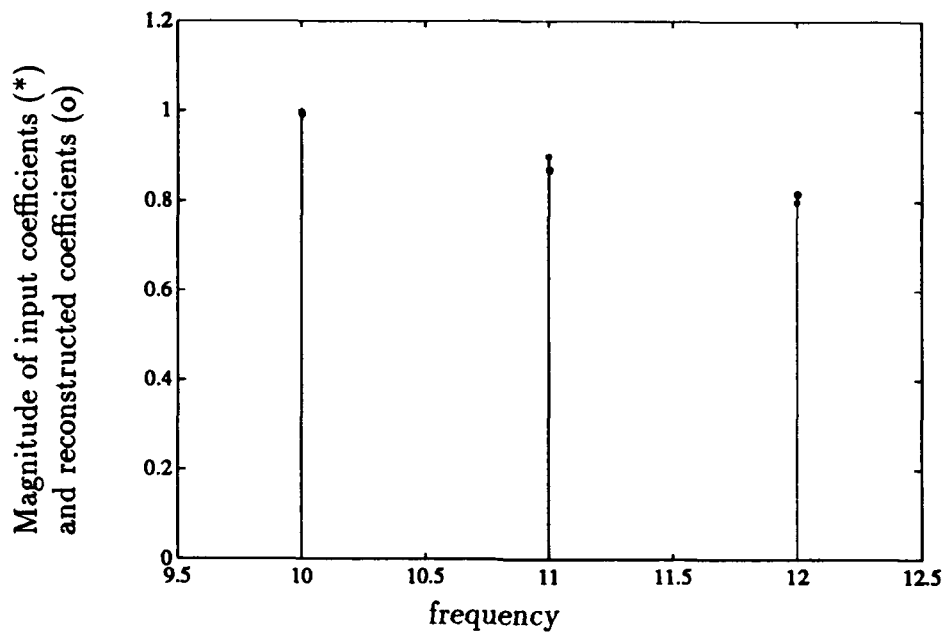
(5) The pure mathematician and the applied mathematician will each have a further question. The pure mathematician may wonder if all the 'sufficiently large' and 'negligible' may not hide some basic flaw in the argument. In my view the arguments above are rigorous but future work will be devoted to a formally stated and proved theorem. The applied mathematician is aware that 'order of' statements may hide impractically large constants. Our answer is essentially 'Suck it and see' (though you should wait until we complete our follow-on work alluded to above). Simple numerical experimentation on our Sun 3/60 suggests that the method is impractical without a modern desktop computer, but that once a speed of 10^6 operations per second is available realistic problems can be tackled.

4 Examples

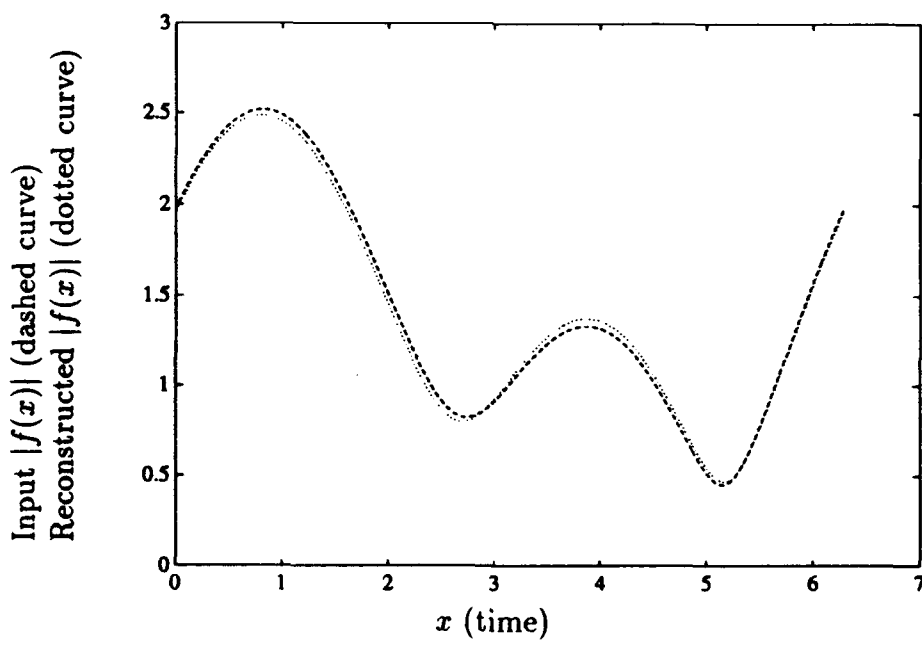
In this section we illustrate our method by demonstrating several examples. Lacking real data with which to exercise the algorithm, we input random values of an "unknown" exponential sum (i.e., known to us, of course, but unknown to the algorithm).

Thus, referring to equation (1), we input the frequencies $\lambda_1, \dots, \lambda_M$, and the coefficients a_1, \dots, a_M . We then choose the parameters and the number of iterations, and turn the algorithm loose on the data. The results are illustrated in the following graphs. Considering the accuracy achieved and the speed with which it is done, on our relatively slow (3 MIPS) Sun 3/60, they speak for themselves.

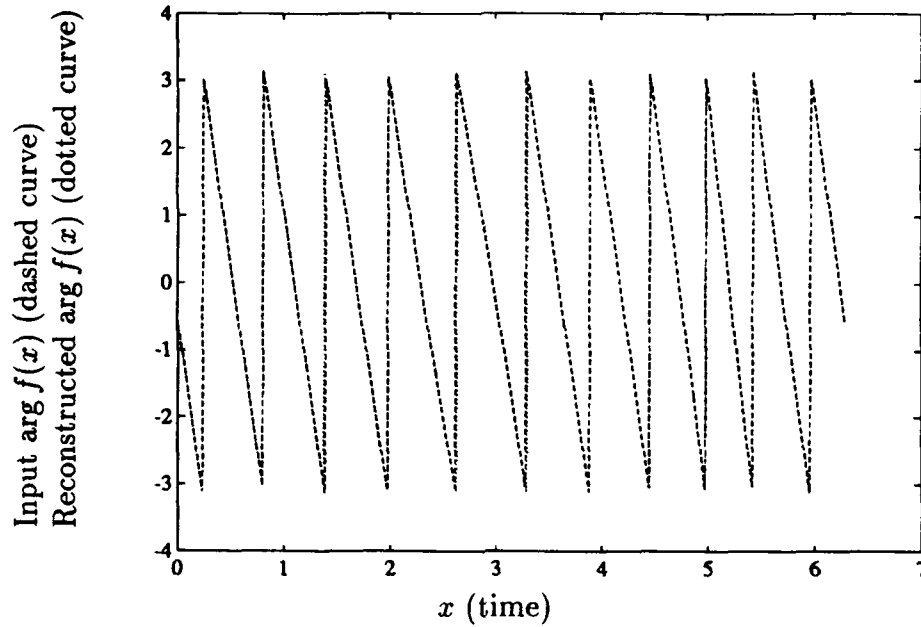
Example 1a



Example 1b



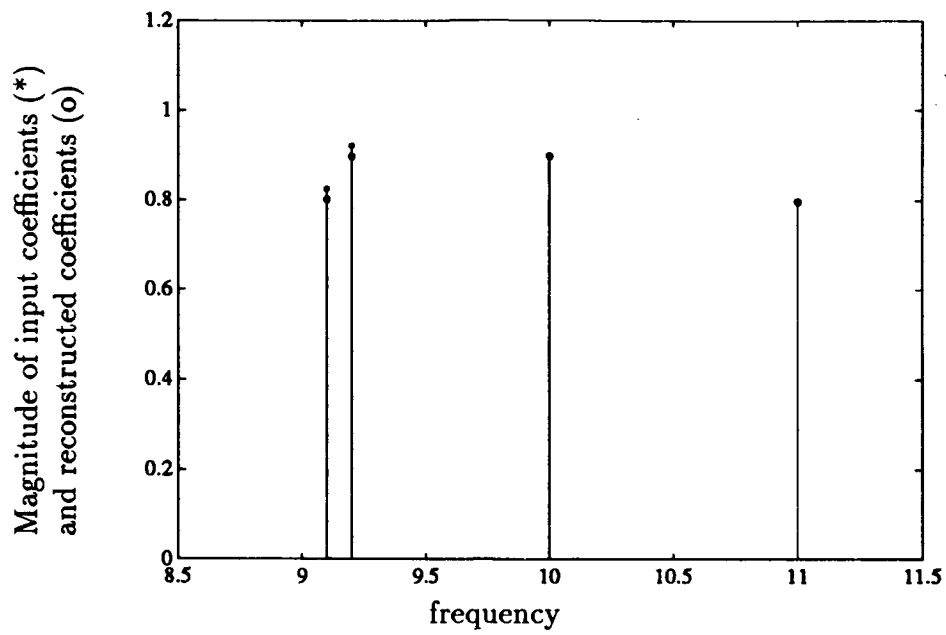
Example 1c



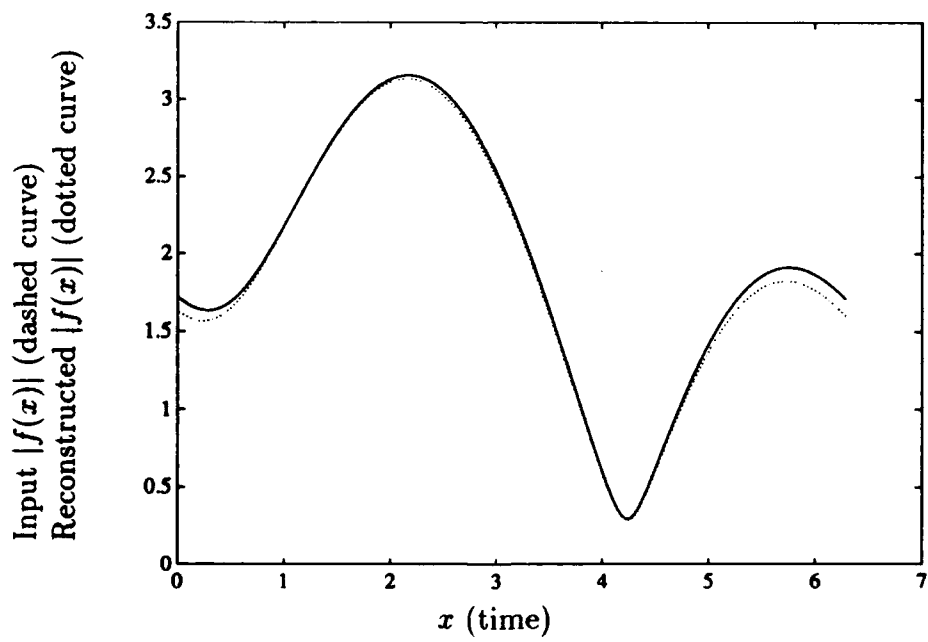
Example 1 parameters

input frequencies = [10, 11, 12]
 input coefficients = [i , 0.9, 0.8]
 noise threshold = 0.6
 samples/iteration = 400
 frequency search resolution/iteration = 100
 number of iterations = 2
 range of frequencies to search = [-20, 20]
 approximate running time = 18 seconds
 reconstructed frequencies = [10.0000, 11.0000, 12.0000]
 reconstructed coefficients = [-0.0138 + 0.9926 i , 0.8689 - 0.0077 i , 0.8162 + 0.0255 i]

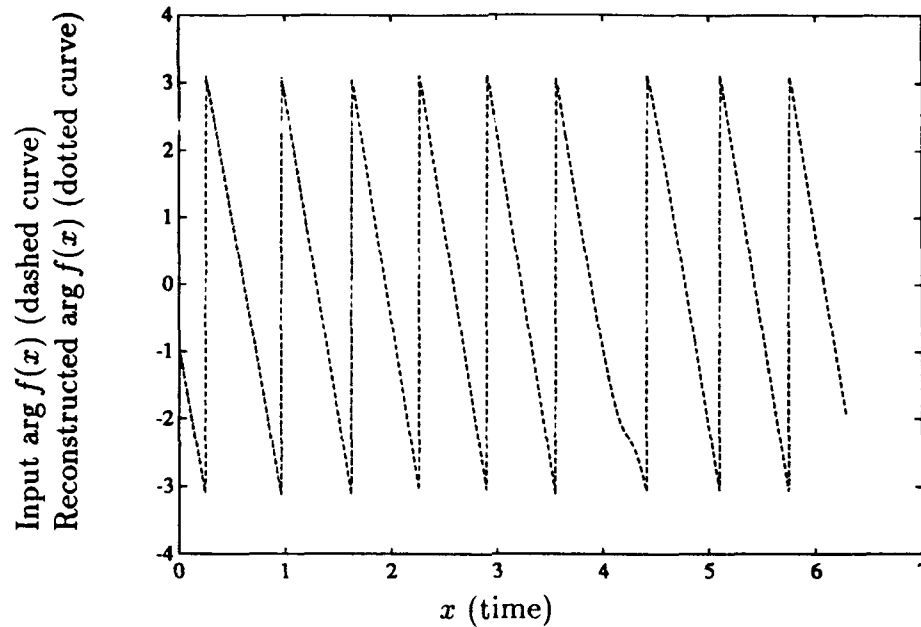
Example 2a



Example 2b



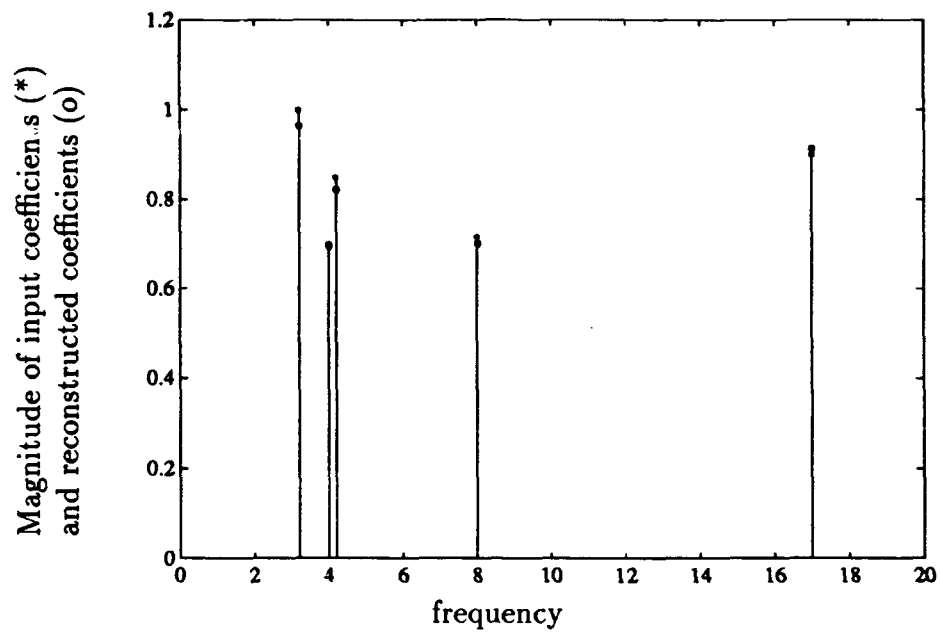
Example 2c



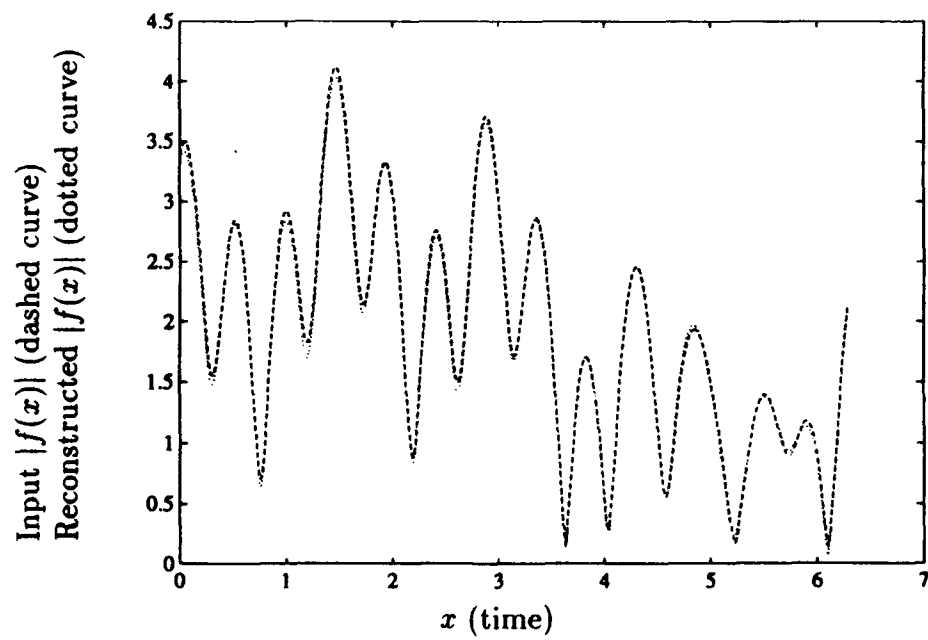
Example 2 parameters

input frequencies = [9.1000, 9.2000, 10.0000, 11.0000]
 input coefficients = [0.2000 + 0.8000i, 0.7000 + 0.6000i, 0.9000, -0.8000]
 noise threshold = 0.6
 samples/iteration = 400
 frequency search resolution/iteration = 100
 number of iterations = 3
 range of frequencies to search = [-20, 20]
 approximate running time = 49 seconds
 reconstructed frequencies = [9.1000, 9.2000, 10.0000, 11.0000]
 reconstructed coefficients = [0.1861 + 0.7981i, 0.6913 + 0.5997i, 0.8977 - 0.0038i,
 -0.7693 + 0.0270i]

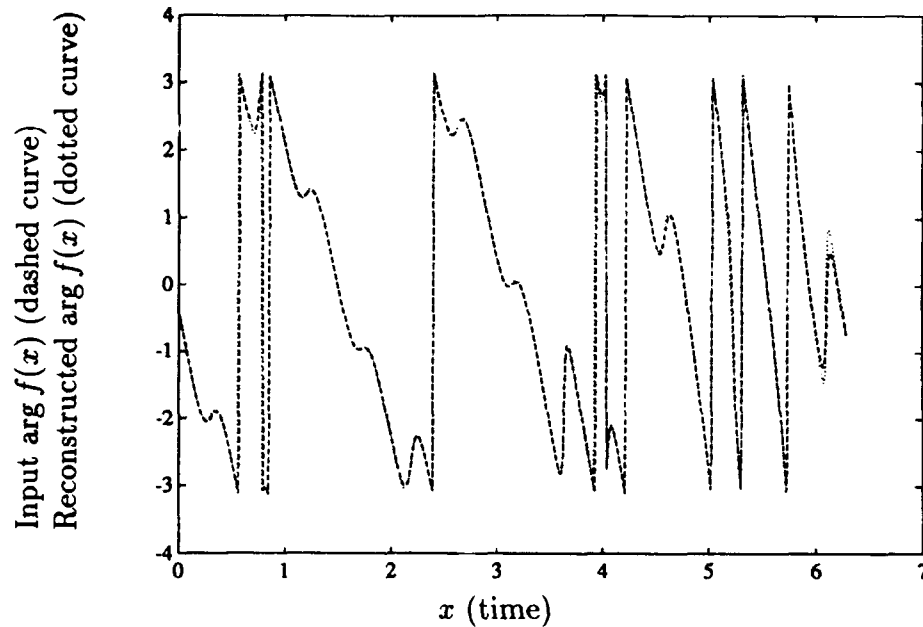
Example 3a



Example 3b



Example 3c



Example 3 parameters

input frequencies = [3.2000, 4.0000, 4.2000, 8.0000, 17.0000]
 input coefficients = [0 + 1.0000i, 0.7000, 0.8500, 0.6500 + 0.3000i, 0.9000]
 noise threshold = 0.6
 samples/iteration = 600
 frequency search resolution/iteration = 150
 number of iterations = 3
 range of frequencies to search = [-20, 20]
 approximate running time = 108 seconds
 reconstructed frequencies = [3.2000, 4.0000, 4.2000, 8.0000, 17.0000]
 reconstructed coefficients = [-0.0285 + 0.9624i, 0.6945 + 0.0003i, 0.8202 - 0.0066i,
 0.6446 + 0.2729i, 0.9131 + 0.0284i]

References

- [BM75] M. Van Blaricum and R. Mittra. A technique for extracting the poles and residues of a system directly from its transient response. *IEEE Trans. Antennas Propagation*, 23:777-781, 1975.
- [HS89] Y. Hua and T.K. Sarkar. A discussion of E-pulse method and Prony's method for radar target resonance retrieval from scattered field. *IEEE Trans. Antennas Propagation*, 37(7):944-946, 1989.
- [JSW83] V.K. Jain, T.K. Sarkar, and D.D. Weiner. Rational modeling by the pencil-of-function method. *IEEE Trans. Acoustics, Speech, Signal Processing*, 31:564-573, 1983.
- [KM65] E.M. Kennaugh and D.L. Moffat. Transient and impulse response approximations. *Proc. IEEE*, 53:893-901, 1965.
- [Per75] A.J. Perni. Target identification by natural resonant estimation. *IEEE Trans. Aerospace Electronic Systems*, 11(2):147-154, 1975.
- [RCN87] E.J. Rothwell, K.M. Chen, and D.P. Nyquist. Extraction of the natural frequencies of a radar target from a measured response using E-pulse techniques. *IEEE Trans. Antennas Propagation*, 35(6):715-720, 1987.
- [S⁺80] T.K. Sarkar et al. Suboptimal approximation/identification of transient waveform from electromagnetic systems by pencil-of-functions method. *IEEE Trans. Antennas Propagation*, 28(6):928-932, 1980.

Prometheus Inc.
Final Report
26 November 1991

Section VI
A computationally efficient notch filter

A computationally efficient notch filter†

James S. Byrnes

Prometheus Inc.

21 Arnold Avenue

Newport, RI 02840

USA

`jbyrnes@cs.umb.edu`

Also at University of Massachusetts at Boston.

1. Introduction

In real-time applications, a digital filter designer may be willing to sacrifice some aspects of performance in return for a significant reduction in computations. A finite impulse response FIR (but not linear phase) notch filter design, in which the number of coefficient multiplications is less than half of the square root of the order N of the filter, is presented here. This expands upon previous work of the author [Byr88]. The trade-off is a considerable passband ripple and a wider notch versus a great computational savings over other means of notch filtering. The deviation from a flat passband fluctuates at a high rate across frequency, and thus may not present a difficulty where the subsequent signal processing is relatively wide band. On the other hand, the notch is somewhat wider than that which may be achieved using window design techniques.

The development arose out of a classic problem in complex analysis in which it is desired to find polynomials, having coefficients of unit magnitude, with nearly a constant magnitude when evaluated along the unit circle [Lit61]. Such a polynomial may be constructed by concatenating the M basis functions of an M -dimensional discrete Fourier

† Research sponsored by the Air Force Office of Scientific Research (AFSC), under Contracts F49620-88-C-0028 and F49620-90-C-0023.

transform DFT into an M^2 -length sequence [Byr77]. The magnitude of the z -transform of the sequence, along the unit circle, oscillates about the value M , taking on the precise value at M^2 equally spaced frequencies.

A notch filter, with the notch at zero frequency, is obtained by leaving out the constant basis function, and reducing the dimension of the filter to $M^2 - M$. Multiple notches may be introduced by setting to zero other basis functions in the sequence. The magnitude frequency response of an example multiple notch filter is shown in Figure 1. The two notches are at zero frequency and a normalized frequency of 0.625 Hz. The filter coefficients are complex-valued and therefore the frequency response is not symmetric. The response values for both positive and negative frequency complex exponentials are shown.

As a consequence of the frequency shifting theorem of the z -transform, a single notch can be moved to any other frequency by multiplying the $M^2 - M$ element coefficient sequence by the corresponding complex exponential frequency. The general class of such filters are referred to here as *constant modulus filters* (CMF) because the magnitudes of the (nonzero) coefficients are equal, while the phase angles vary.

The reduction in coefficient multiplications is due to the repeating DFT elements. For an M -dimensional DFT there are only M distinct factors in an array of M^2 factors of the basis functions. The M factors are repeated in a pattern of varying order, with some occurring more frequently than others. In a direct-form, or transversal filter realization of the notch filter, the input sequence elements which are to be multiplied by the same factor may be added together before the coefficient multiplication. Further reduction in coefficient multiplications by a factor of two is achieved if M is an even number. When M is even, each element in the sequence is accompanied by the negative value

of the same element. Thus the corresponding elements in the input sequence may be subtracted before the coefficient multiplication.

Table 1.1 shows the number of coefficient multiplications per time step for $M = 4, 6, 8, 10$ and 12 . Multiplication by 1 or by powers of the imaginary unit i ($90, 180$ and 270 degree phase shifts) are not counted because they can be implemented by adding and subtracting.

<i>filter</i>	M^2	M	<i>number of actual coeff.</i>
<i>dimension</i>			<i>multiplications</i>
12	16	4	none
30	36	6	2
56	64	8	2
90	100	10	4
136	144	12	4

Table 1.1: Coefficient multiplications per time step

In section II the magnitude frequency responses of the filters listed in Table 1.1 are compared to those which can be obtained using a Hamming window of similar dimension. Section III presents the result of gross rounding of the remaining fractional coefficients to further reduce the computational burden.

2. Frequency responses

Figures 2a through 2e are the magnitude frequency responses of the five filters of Table 1.1. For comparison five notch filters, designed using Hamming windows, are also shown. The "Hamming notch filter" coefficients were obtained using odd dimensional

Hamming windows. An odd number of elements is required for such a design. All of the filter coefficients except the center coefficient are made proportional to the negative of a Hamming window. The center element is adjusted so that the average of the filter coefficients is equal to zero.

The notch filters designed using Hamming windows do not yield the most narrow notches which could be compared to the constant modulus filters, but they are used here as a standard of comparison which is familiar to the reader. Indeed, if passband ripple is of little consideration, a notch filter designed using a rectangular window will yield a more narrow notch for a given dimension than any of the classical windows.

The filter magnitude frequency responses are not symmetric functions, as is the case with real-valued filter coefficients, so they are shown over the entire range of normalized frequency, from zero to the sample rate, with the sample rate normalized to 1 Hz. A surprising feature of the set of frequency responses shown is that the peak-to-peak variation in response does not change significantly with the filter order. In each case the maximum passband variation is approximately -20% to $+25\%$; there is a slight improvement with increase in filter order.

Whereas the width of the notch produced by the Hamming filter decreases in inverse proportion to the dimension of the filter, the CMF do behave so, but vary from a factor of two to a factor of five times the width of the corresponding Hamming notch filter.

Figures 3a through 3d show the frequency responses on a log-log basis for the 30-, 56-, 90-, and 132-dimensional CMF filters in comparison to the 31-, 57-, 91-, and 133-dimensional Hamming notch filters. Only the negative exponential frequencies are shown. These correspond to the right hand side of the corresponding curves in Figure 2. Each of the Hamming notch filters has a double zero at 1, and therefore the curves approach

a slope of 40 dB/decade at the lower frequencies. The CMFs, on the other hand, have only a single zero at 1; these approach a slope of 20 dB/decade at lower frequencies.

The width of the notch is more easily read from the log-log plots than from the linear frequency response plots. If the *width* is considered to mean twice the distance from zero frequency to the first crossover of 0 dB, or to the point of departure of the frequency response from 0 dB for the Hamming notch filter, the comparative notch widths are shown in Table 2.1. The CMF notch widths will vary slightly depending on the notch frequency. Also, the notches are not symmetric. The numbers shown in the table are from the curves of Figure 3.

<i>filter</i>	<i>CMF notch</i>	<i>Hamming notch</i>	<i>ratio</i>
<i>dimension</i>	<i>width (Hz)</i>	<i>width (Hz)</i>	
30,31	0.130	0.060	2.2
56,57	0.030	0.011	2.7
90,91	0.090	0.020	4.5
132,133	0.068	0.014	4.9

Table 2.1: Half bandwidths of notches in normalized frequency

3. Rounded coefficients

Gross rounding of the filter coefficients was tested for the dimension 30 and 56 filters to see if they would retain their essential properties despite significant change in the remaining two fractional coefficients. For the dimension 30 filter, these are:

$$b(1) = \exp(i\pi/4) \quad (1)$$

$$b(3) = \exp(i3\pi/4) \quad (2)$$

These were adjusted to the values:

$$\tilde{b}(1) = 1 + i \quad (3)$$

$$\tilde{b}(3) = -1 + i \quad (4)$$

so that the actual coefficient multiplication could be performed by two additions. The resulting magnitude frequency response is shown in Figure 4a for comparison to the CMF. Note that the rounded version of the filter was not re-normalized to 0 dB in the passband. Therefore the magnitude frequency response curve of the rounded coefficient filter is above that for the CMF. The variation in the response is approximately the same as that for the CMF, and the notch width has not changed noticeably.

More significant is the fact that the zero at 1 has been preserved after the rounding. This fact follows from the symmetric pattern of rounding in which the average value of the filter coefficient sequence remains zero.

For the 56-dimensional filter, the remaining fractional coefficients are:

$$b(1) = \exp(i\pi/3) \quad (5)$$

$$b(2) = \exp(i2\pi/3) \quad (6)$$

These were adjusted to:

$$\tilde{b}(1) = 1/2 + i \quad (7)$$

$$\tilde{b}(2) = -1/2 + i \quad (8)$$

The real parts are unchanged but the imaginary parts have been rounded from $\sqrt{3}/2$ to 1. The magnitude frequency responses of the rounded and CMF versions are shown in Figure 4b. Once again, the rounded version has not been renormalized so that the frequency response curve appears above the CMF. The variation in magnitude frequency response has not significantly changed, nor has the width of the notch. Furthermore, the zero at the origin has been preserved in the rounding off of coefficients.

In both of these cases, the notch cannot be moved to an arbitrary frequency while retaining the computational savings of the rounding of the coefficients. But certain

frequency shifts are possible. A shift to the folding frequency, for instance, can be done by sign changes of the rounded coefficients.

4. Conclusion

A very efficient family of FIR notch filters has been described. The filter coefficients consist of M -dimensional DFT basis functions concatenated into $M^2 - M$ element arrays. The repetition of the DFT factors within the impulse response allows for a great reduction in the number of coefficient multiplications. Gross rounding of the real and imaginary parts of the fractional coefficients can be employed without changing the essential features of the frequency response. The filters are expected to be useful in real-time filtering applications where passband ripple may be traded for computational savings.

5. Bibliography

- [Byr77] J.S. Byrnes. On polynomials with coefficients of modulus one. *Bull. London Math. Soc.*, 9:171-176, 1977.
- [Byr88] J.S. Byrnes. A notch filter employing coefficients of equal magnitude. *IEEE Trans. Acoustics, Speech, Signal Processing*, 36(11):1783-1784, 1988.
- [Lit61] J.E. Littlewood. On the mean values of certain trigonometric polynomials. *J. Lon. Math. Soc.*, 36:307-334, 1961.

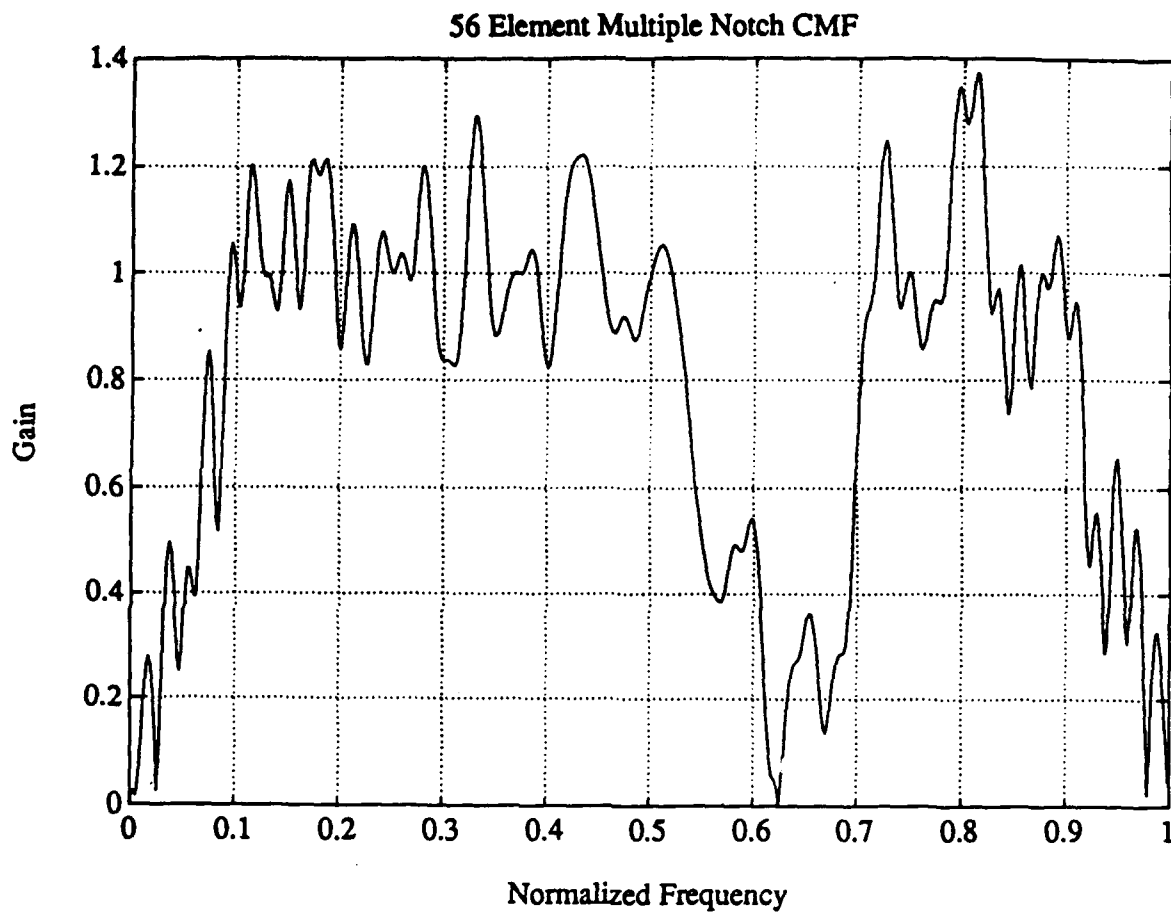


Figure 1

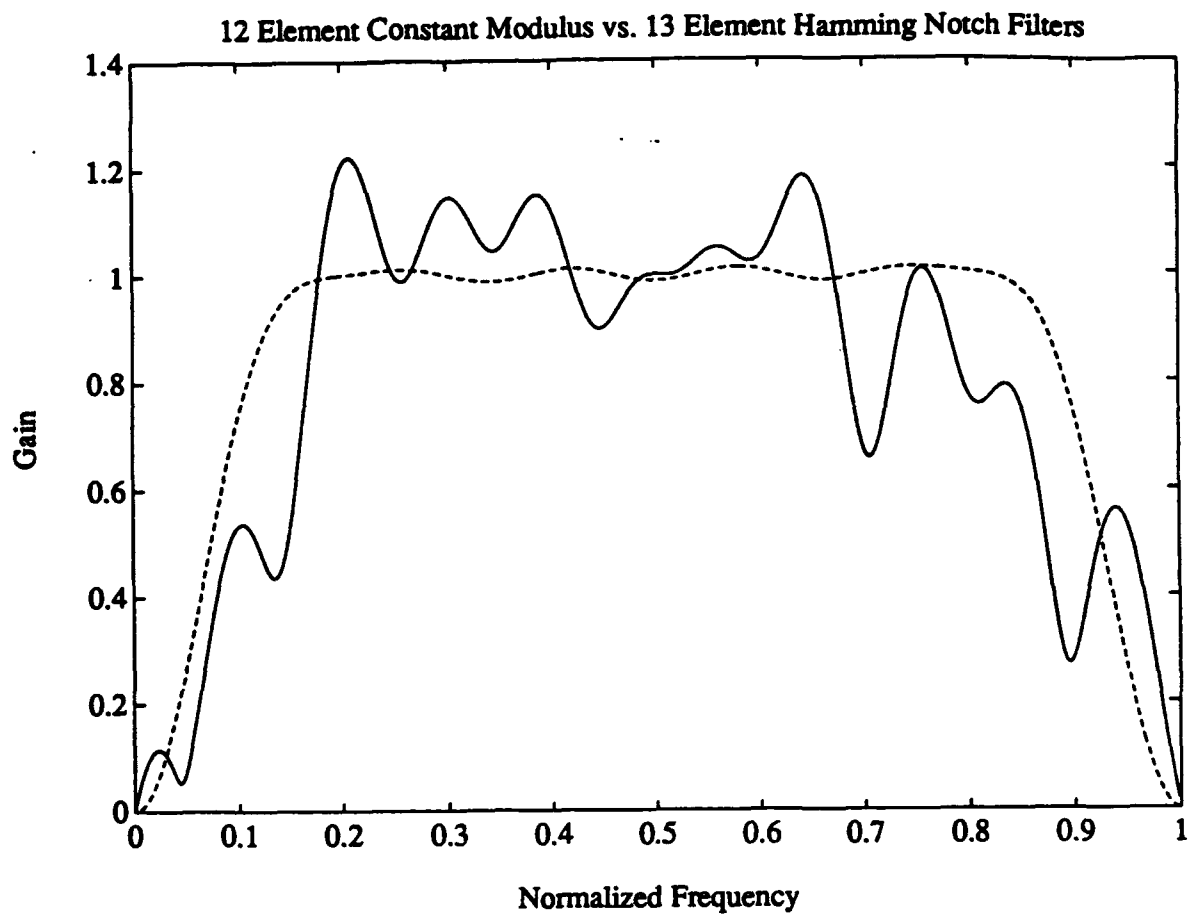


Figure 2a

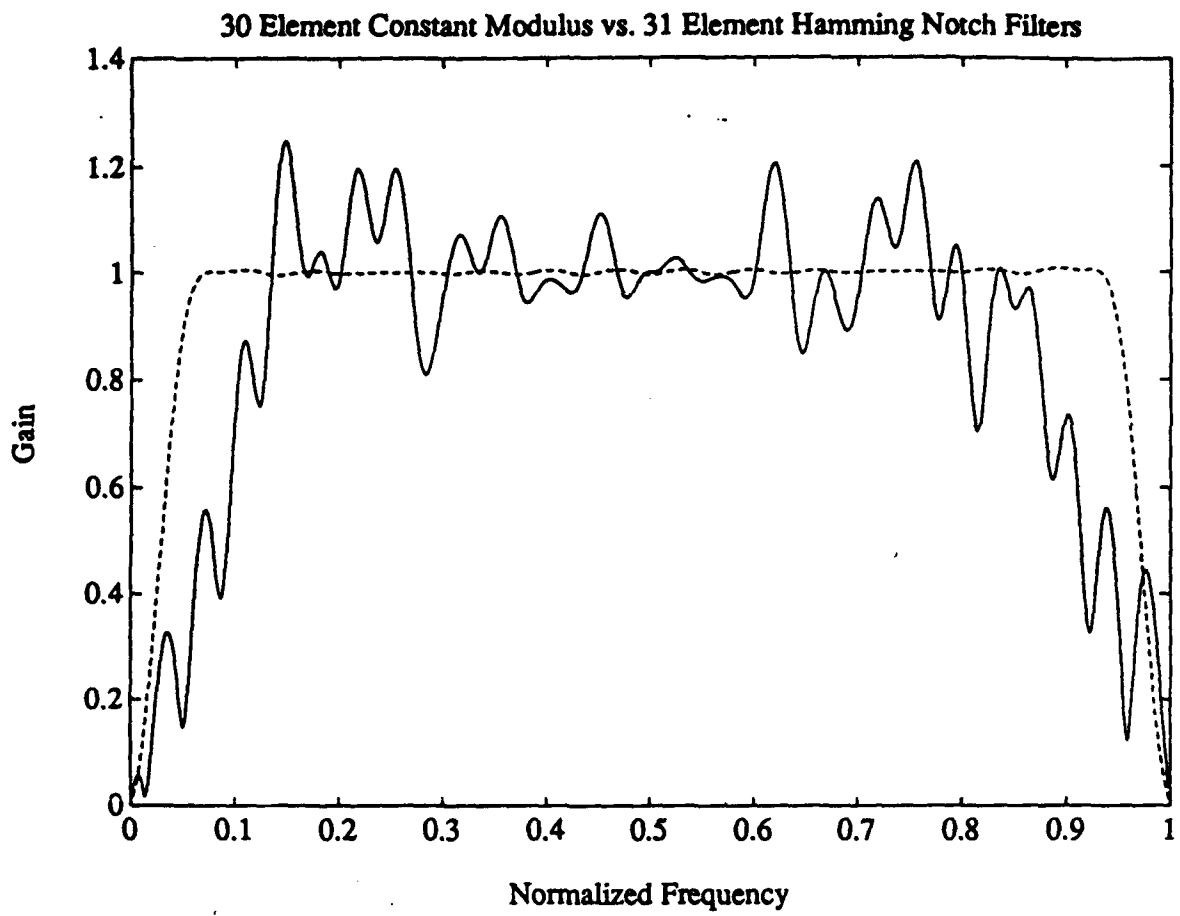


Figure 2b

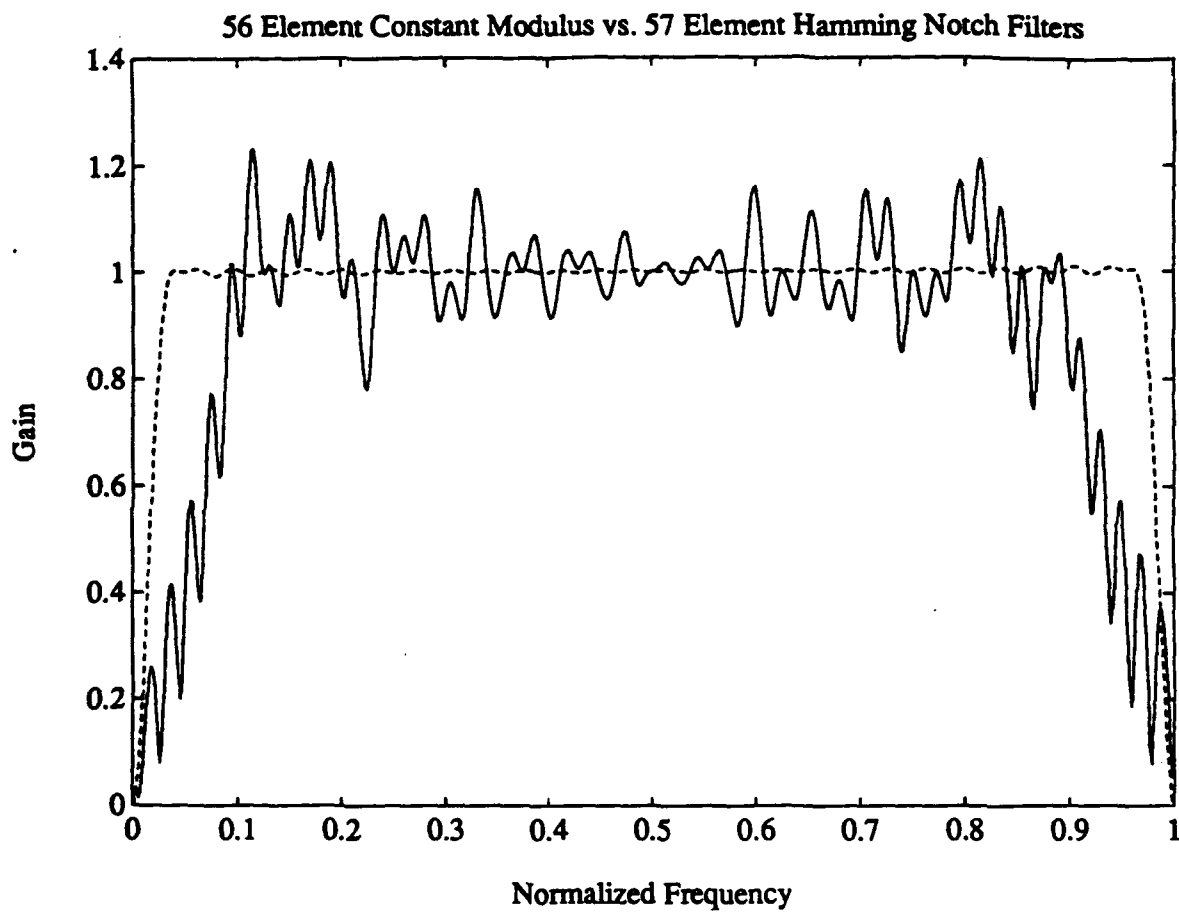


Figure 2c

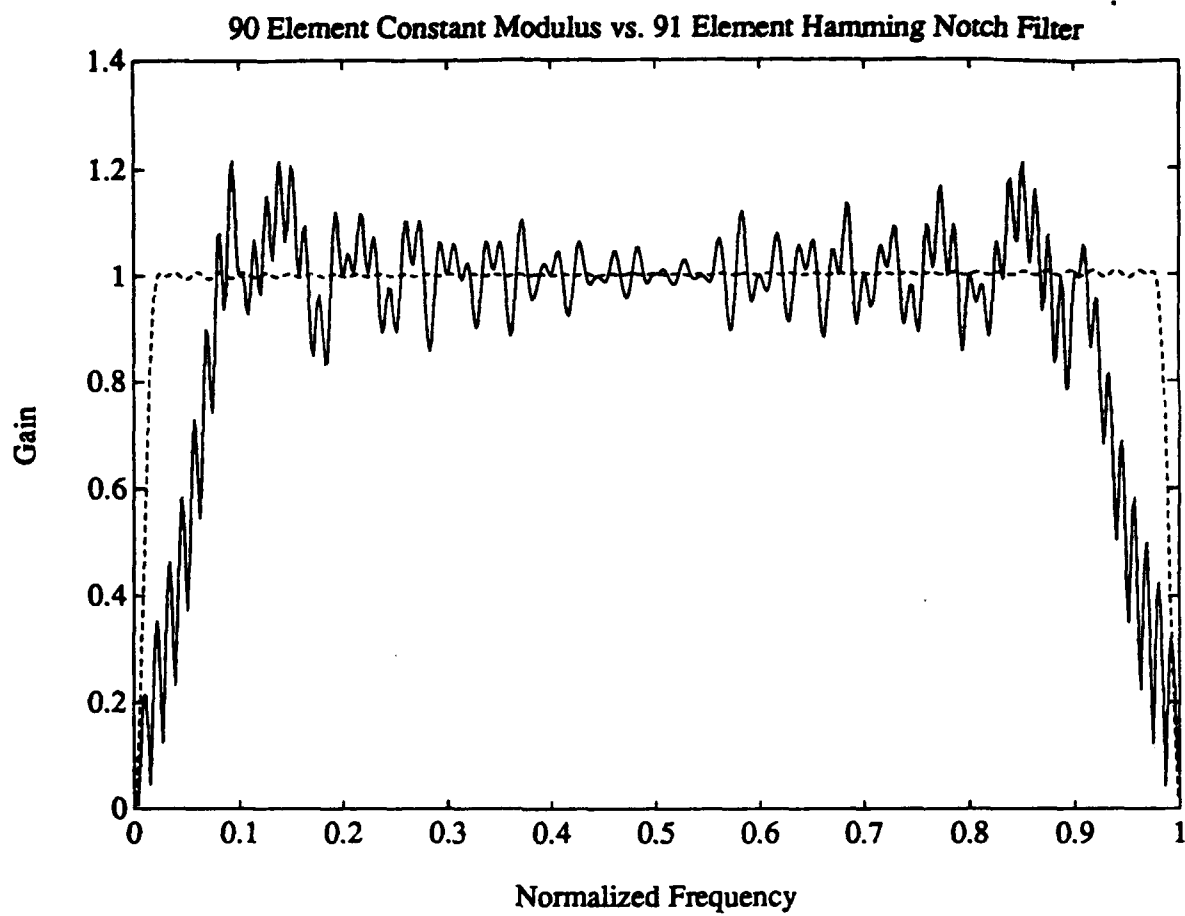


Figure 2d

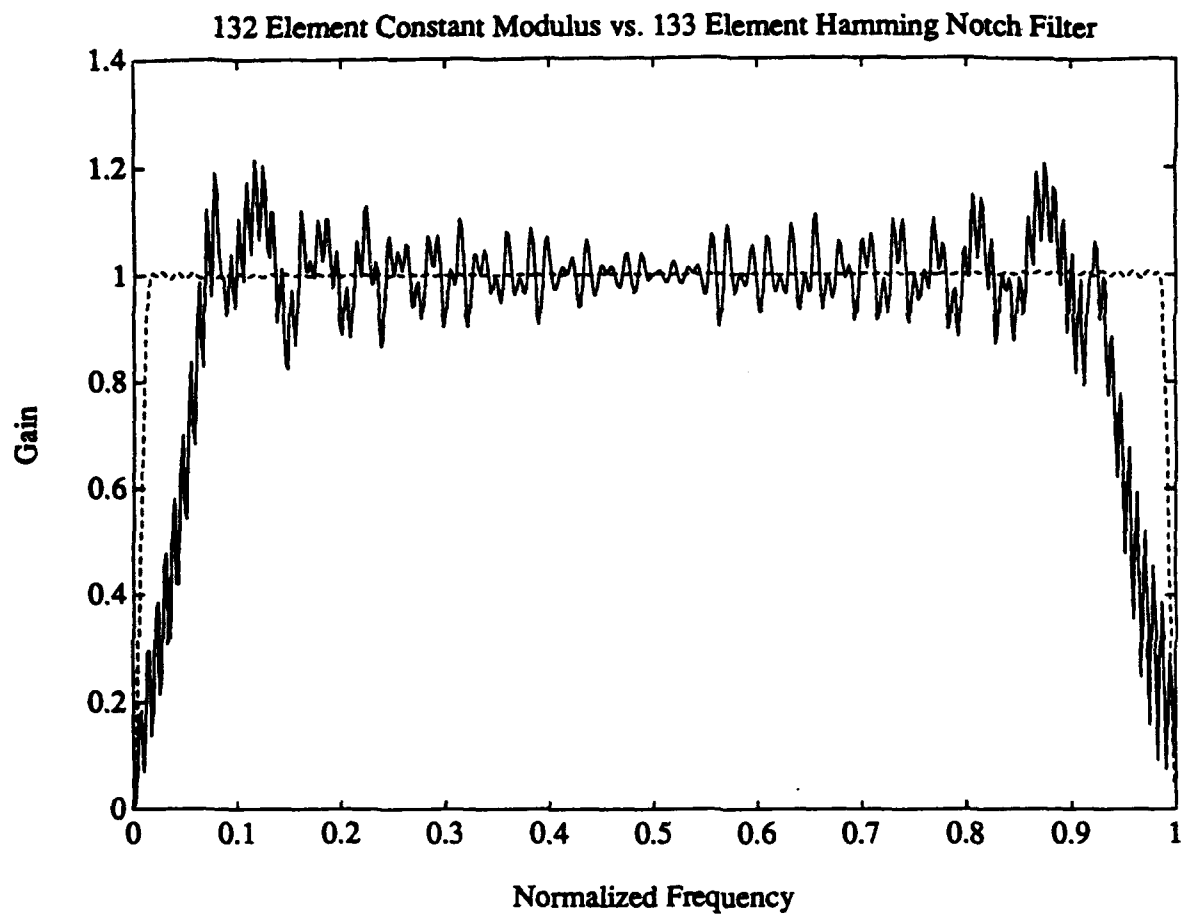


Figure 2e

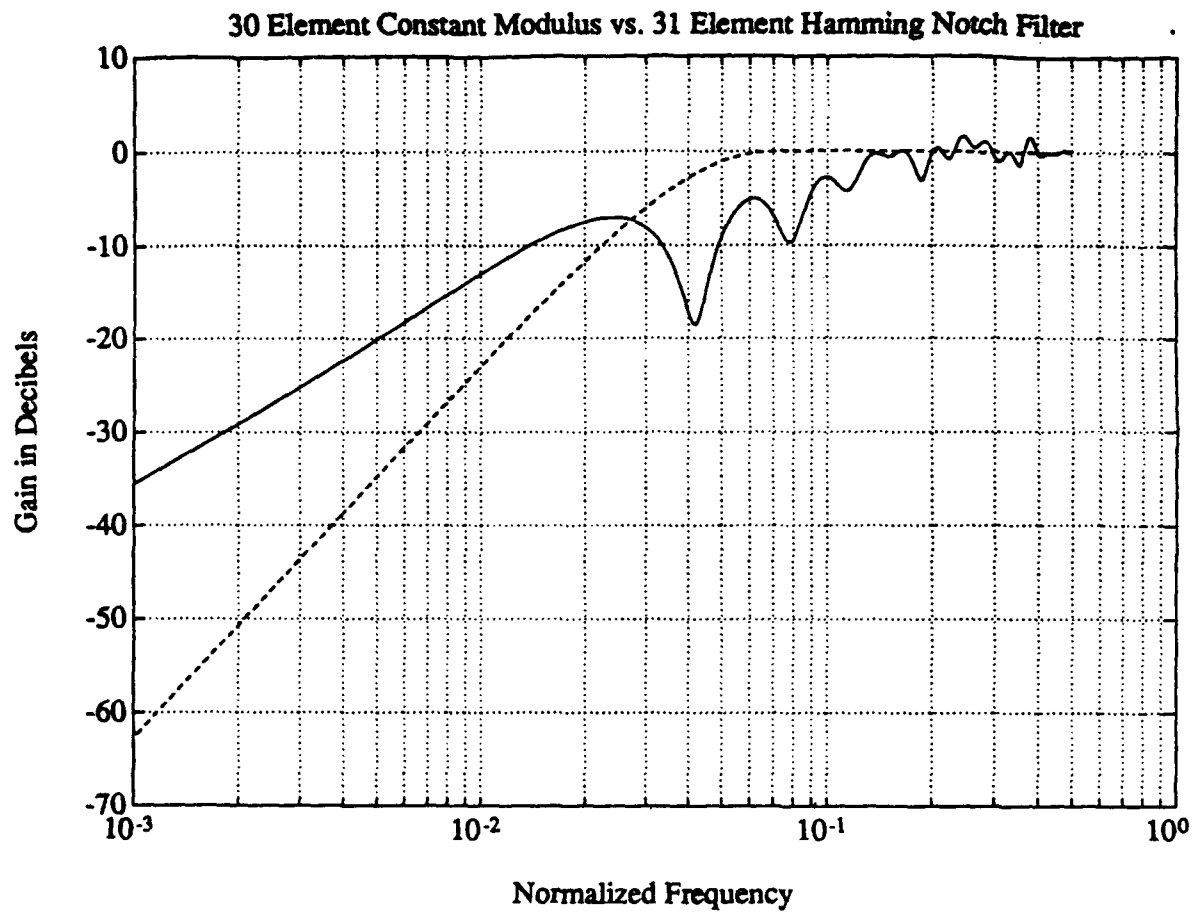


Figure 3a

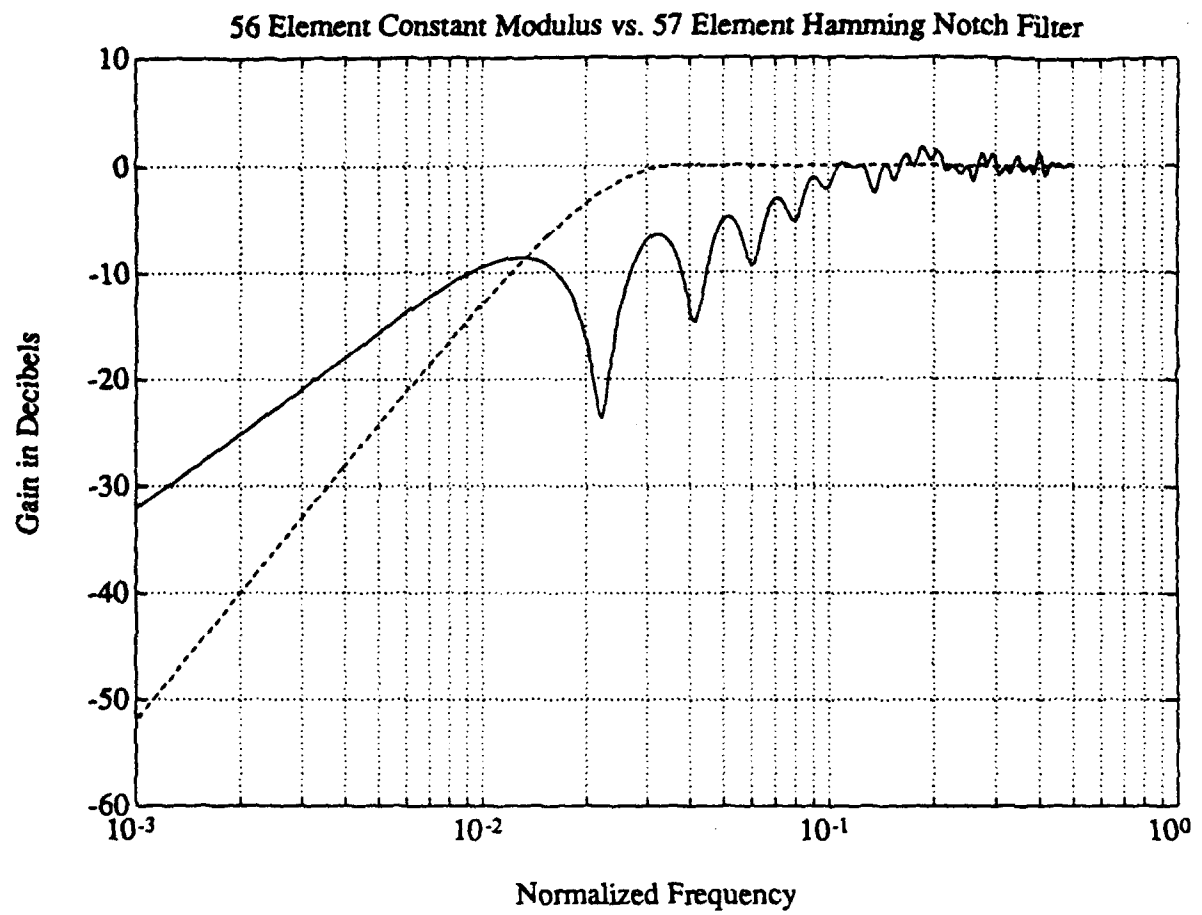


Figure 3b

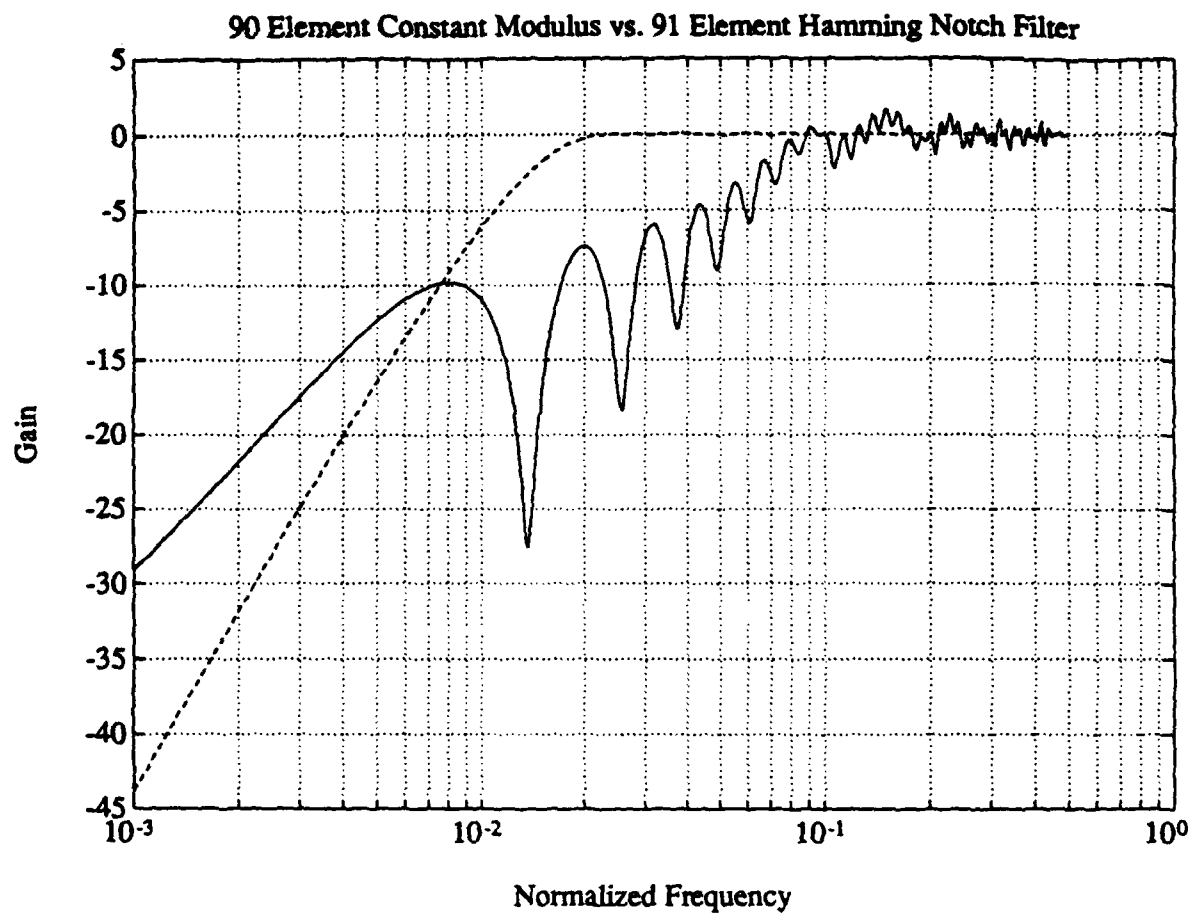


Figure 3c

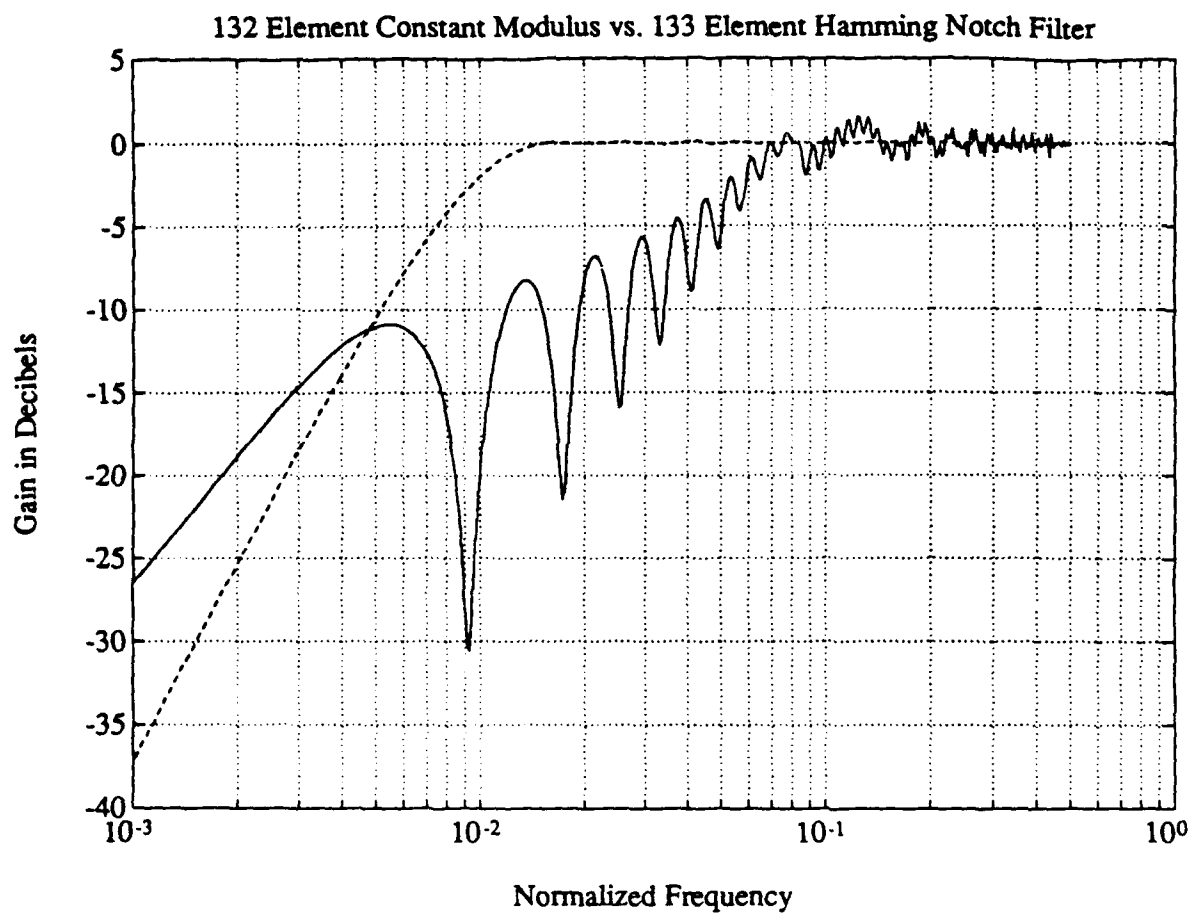


Figure 3d

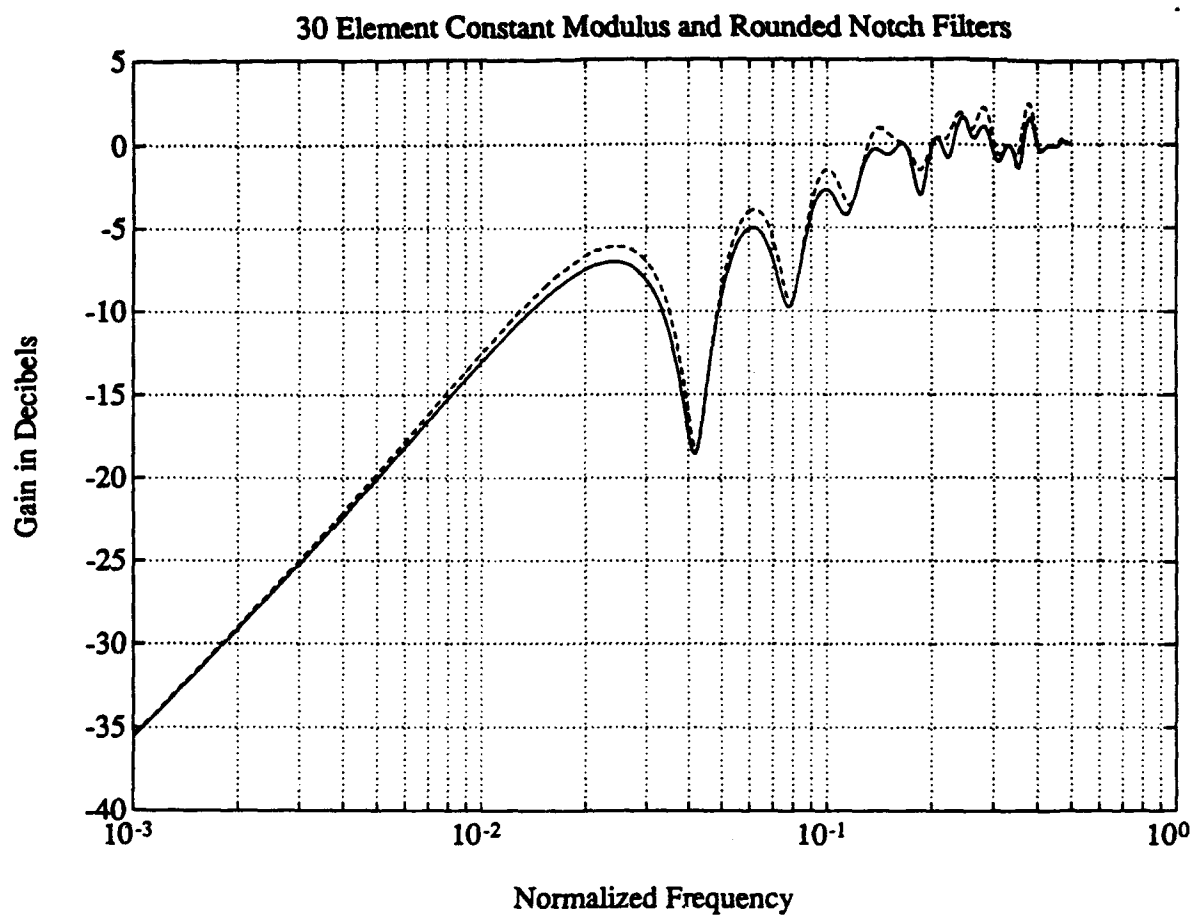


Figure 4a

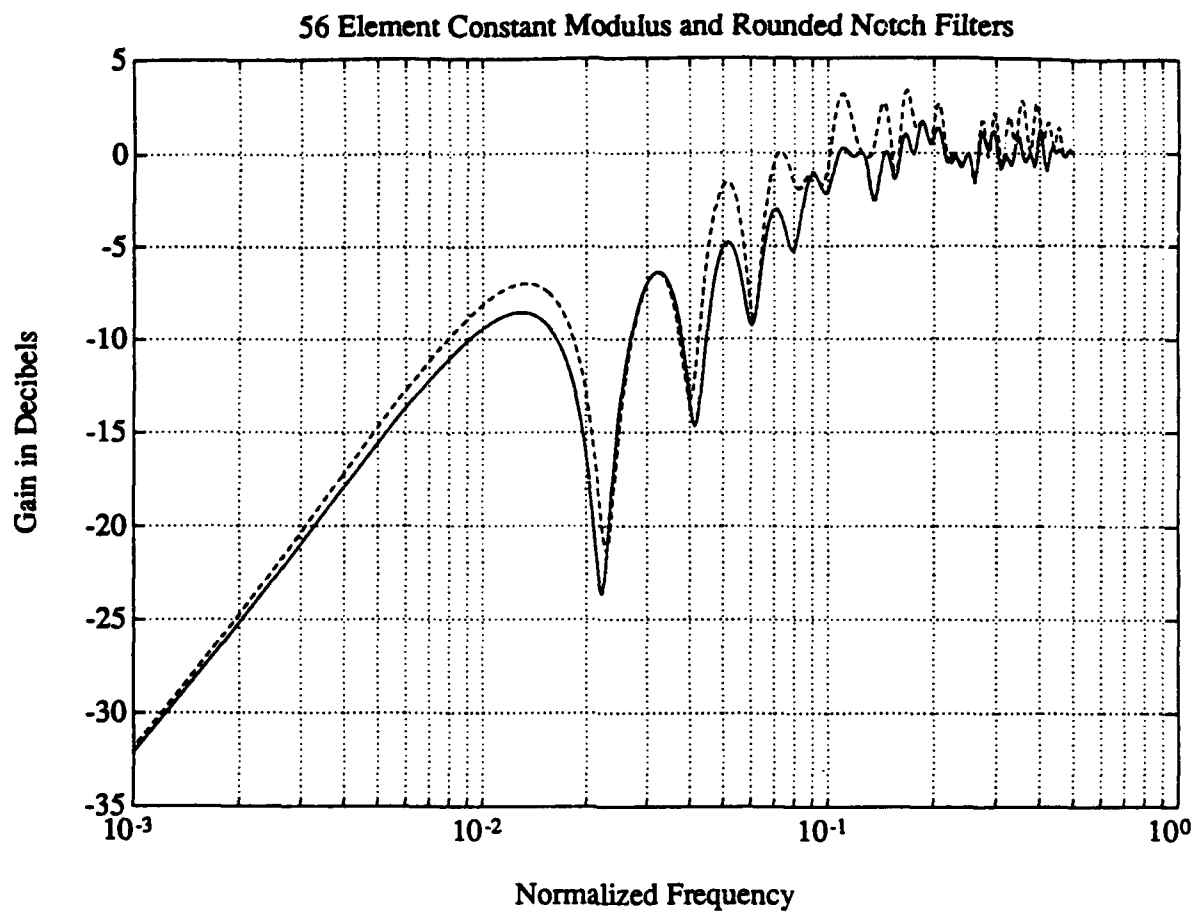


Figure 4b

Prometheus Inc.
Final Report
26 November 1991

Section VII
A new rational approximation to digital filters

A New Rational Approximation to Digital Filters*

Donald J. Newman
Prometheus Inc.
21 Arnold Avenue
Newport, RI 02840
Also at Temple University.

Gerald Ostheimer
Prometheus Inc.
Also at St. Andrews University.

J.S. Byrnes
Prometheus Inc.
Also at University of Massachusetts at Boston.

December 21, 1990

1 Introduction

A key electrical engineering application of approximation theory occurs in the design of recursive digital filters. One ordinarily wants a preassigned frequency selective behavior, requiring consideration of both the magnitude of the filter frequency response and the desired characteristic which is to be approximated. Thus, in a typical lowpass filter approximation problem, one is given the passband, the stopband, the maximum attenuation in the passband,

*Research supported by AFOSR, under contract F49620-90-C-0023

and the minimum attenuation in the stopband. The resulting mathematical problem is to choose a rational function of appropriate degree, with correctly placed poles and zeroes, which approximates the desired frequency response according to the specified passband and stopband parameters. Vlček and Unbehauen [VU89], expanding upon the work of Daniels [Dan74], offer one analytic approach to this problem for the case of an IIR filter with equiripple magnitude behavior.

More generally, the digital filter design problem requires the approximation of a (real-valued) filter amplitude characteristic, or a (complex-valued) filter characteristic, by a realizable digital filter transfer function. This transfer function is a (real or complex-valued, respectively) rational function of a complex variable. Furthermore, in addition to being a good approximation in some norm, it must also satisfy certain constraints, such as a bound on its degree, the location of its poles, etc. Thus we are led, in many applications, to the following type of minimization problem [CSS77].

Let

$$H(z^{-1}) = \sum_{n=-\infty}^{\infty} a_n z^{-n}, \quad z = e^{i\theta} \quad (1)$$

(the "z transform," in engineering terms) be an L^2 function on the unit circle in the complex plane; let R be some class of rational functions of z ; and let $\|\cdot\|$ be a norm (or semi-norm). Study the existence, uniqueness (local or global), characterization, etc., of an $r^* \in R$ satisfying

$$\|H - r^*\| = \inf \{ \|H - r\| : r \in R \}. \quad (2)$$

The existence and uniqueness of such a best approximation, when $\|\cdot\|$ is the L^2 norm, the degrees of both the numerator and denominator of r are fixed, and the poles of r are given, is shown in [CSS77]. They also produce an upper bound on the degree of approximation in this case, when the poles are equally spaced on a circle about the origin and the degree of the numerator is restricted to be equal to that of the denominator.

On the other hand, one may also approximate the one-sided z-transform

$$\tilde{H}(z^{-1}) = \sum_{n=0}^{\infty} \tilde{a}_n z^{-n}$$

of the sample sequence $\{\tilde{a}_n\}$ by matching the coefficients of the Maclaurin expansion of the rational function with the response samples $\{\tilde{a}_n\}$ as far as

possible. This, of course, leads to a Padé approximant of $H(z^{-1})$. This technique has been used, but with limited success, in the synthesis of recursive digital filters ([BS73], [MG83]).

These results are certainly of theoretical interest, but clearly for practical problems much remains to be done. For instance, there is no acceptable algorithm for the approximation of an arbitrary frequency response or magnitude characteristic by a stable IIR filter in the Chebyshev norm. In the frequency response case, there is no algorithm which, beginning from an arbitrary starting point, converges to an optimal solution. The existing near-best approximation techniques are often unstable.

2 Discussion

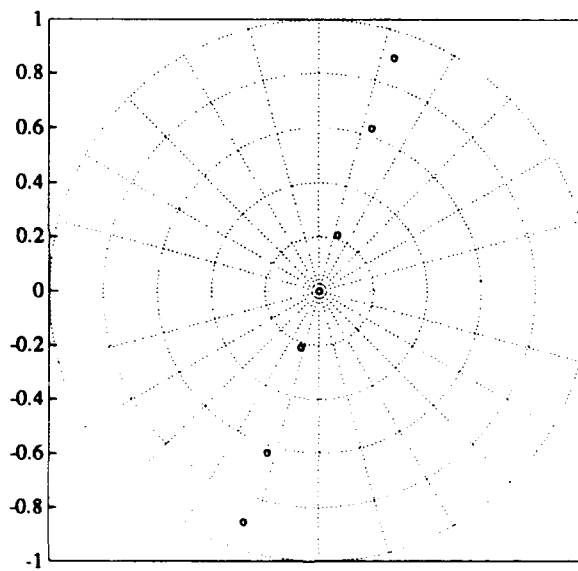
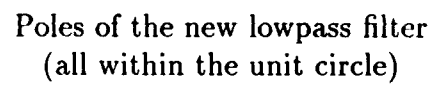
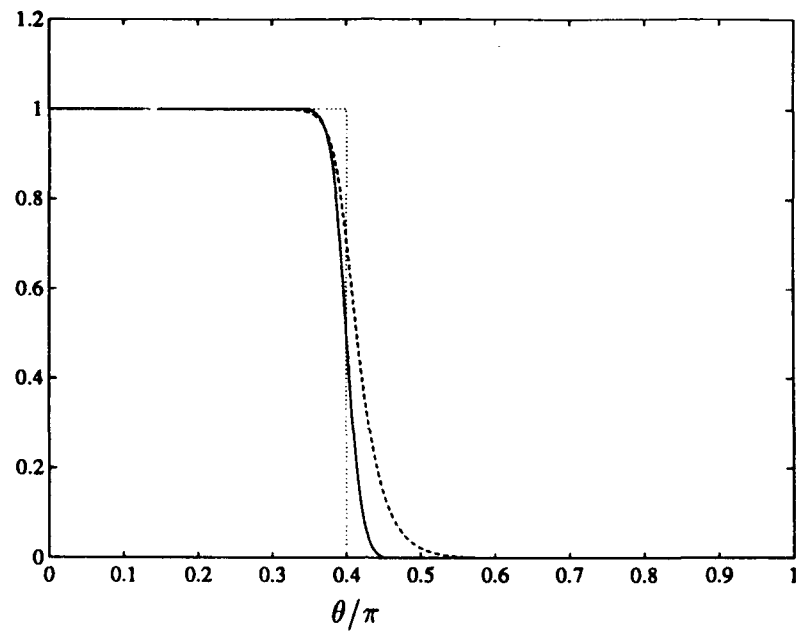
In this note we present a new method of approximating a digital filter by a rational function.

Our approach is to base the required rational approximation upon the seminal paper [New64] of the first author. As this work was the first to demonstrate a qualitative difference between rational and polynomial approximations, it is considered fundamental in the field. Once the concepts of this paper are understood, their application to the construction of digital filters is rather straightforward. Essentially all that is required is a change of variable to transform the real interval $[-1, 1]$ to the unit circle, a scaling, an inversion of the poles outside the unit circle to move them inside, and a further change of variables to adjust the passband. We illustrate the results with two examples, done in MATLAB on a Sun 3/60.

For purposes of comparison, we also illustrate the standard Butterworth filter, as its requirements and performance are closest to the new filter described here.

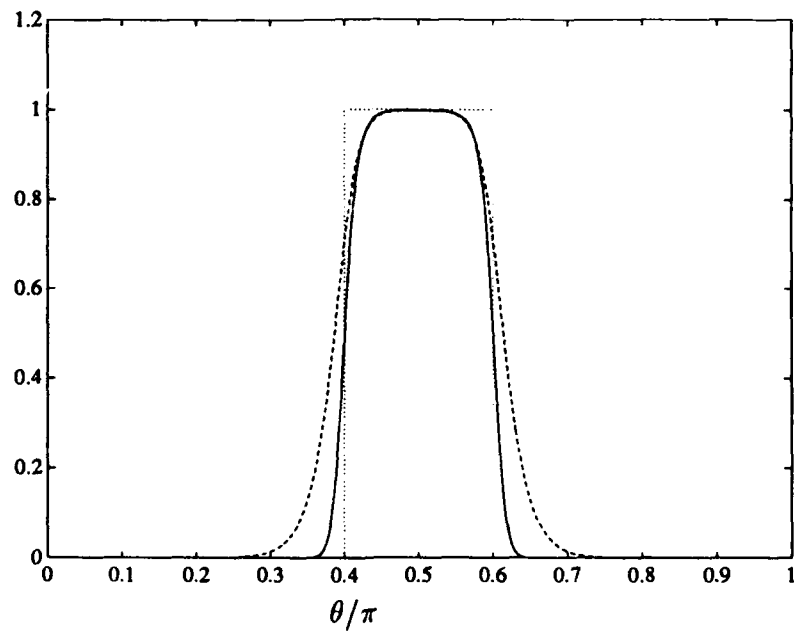
3 Examples

Example 1: lowpass filter

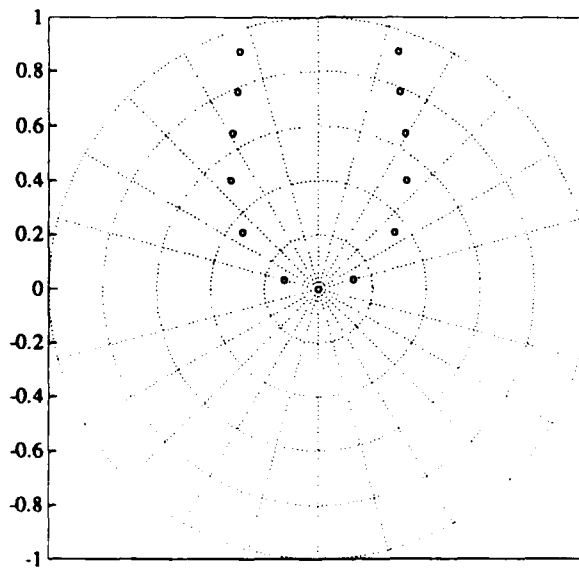


Ideal bandpass filter (dotted curve)
 Butterworth bandpass filter, degree 24 (dashed curve)
 New bandpass filter, degree 24 (solid curve)

Example 2: bandpass filter



Poles of the new bandpass filter
 (all within the unit circle)



References

- [BS73] F. Brophy and A.C. Salazar. Consideration of the padé approximant technique in the synthesis of recursive digital filters. *IEEE Trans. Audio Electroacoustics*, 21:500–505, 1973.
- [CSS77] C.K. Chui, P.W. Smith, and L.Y. Su. A minimization problem related to padé synthesis of recursive digital filters. In E.B. Saff and R.S. Varga, editors, *Padé and Rational Approximation*. Academic Press, New York, 1977.
- [Dan74] R.W. Daniels. *Approximation Methods for Electronics Filter Design*. McGraw-Hill, New York, 1974.
- [MG83] Edmund K. Miller and Dennis M. Goodman. A pole-zero modeling approach to linear array synthesis 1. the unconstrained solution. *Radio Science*, 18(1):57–69, 1983.
- [New64] D.J. Newman. Rational approximation to $|x|$. *Michigan Mathematical Journal*, 11:11–14, 1964.
- [VU89] M. Vlček and R. Unbehauen. Analytic solutions for design of iir equiripple filters. *IEEE Trans. Acoustics, Speech, Signal Processing*, 37(10):1518–1531, 1989.

Prometheus Inc.
Final Report
26 November 1991

Section VIII
Concerning Prony's method

Concerning Prony's Method†

Harold S. Shapiro
Prometheus Inc.
21 Arnold Avenue
Newport, RI 02840

Also at KTH, Stockholm.

J.S. Byrnes
Prometheus Inc.

Also at University of Massachusetts at Boston.

1. Introduction

When the transient scattered field of a radar target is approximated by an exponential sum

$$y(t) = \sum_{i=1}^M b_i e^{s_i t}, \quad 0 \leq t \leq t_y, \quad (\text{notation from [HS89]})$$

the s_i are the dominant resonant poles of the target. As they represent the resonant frequencies of the target, they are ordinarily the most important information available in the transient field. Since these frequencies are aspect independent they give a signature of the object, and thus can be used as a means of target identification [Per75].

The study of efficient methods for estimating the poles of radar targets, apparently initiated by Kennaugh and Moffatt [KM65], remains an active area of research. The techniques employed thus far include Prony's method [BM75], the pencil-of-function approach [S⁺80, JSW83], and the E-pulse method [RCN87, HS89]. Each of these methods have their particular advantages and drawbacks, with none being close to ideal in any global sense. For example, one problem with Prony's method is its reliance on uniformly sampled sinusoidal functions. In this paper we examine Prony's method from a more mathematical viewpoint than is ordinarily taken, and we consider some alternatives.

2. Work

The Problem: We are given a sequence $\{c_0, \dots, c_m\} \in \mathbb{C}$ and we ask: are there complex numbers a_1, \dots, a_r and z_1, \dots, z_r such that

$$c_k = \sum_{j=1}^r a_j z_j^k \quad k = 0, 1, \dots, m? \quad (2.1)$$

† This research was supported by AFOSR under contract F49620-90-C-0023.

Remark. We can think of r as given, $r \geq 1$. Note that if we choose $r = m + 1$, the problem is solvable by taking *any* choice of distinct z_1, \dots, z_{m+1} , and solving (2.1) for the a_j (Van der Monde system).

If $r < m + 1$ then, in general, there is no solution. Thus, take $r = m$ in (2.1). If $c_0 = c_1 = \dots = c_{m-1} = 0$ then all the a_j must vanish, and $c_m = 0$ too. Thus the values $m = 0, 0, \dots, 0, 1$ can't be interpolated by a function of type (2.1). We'll see another "reason" for this shortly. (Remark, p.3)

In principle we want to take r as small as possible for the given $\{c_j\}$. However, for "physical" reasons we may want the $\{z_j\}$ to satisfy extra requirements, e.g., to lie in \mathbb{D} (the unit disk), or $\bar{\mathbb{D}}$, or on the unit circle, or near it, etc. This may motivate a choice of r larger than what, purely algebraically, is required by (2.1). (Also, we may want the $\{a_j\}$ subjected to certain bounds.)

Theorem 2.1 (Prony). This theorem says that "generically", (2.1) always has a solution when m is odd and $r = (m + 1)/2$ (so that the number of available parameters $\{a_1, \dots, a_r, z_1, \dots, z_r\}$ equals the number of "data points" $\{c_0, c_1, \dots, c_{2r-1}\}$). A careful analysis shows that Prony proves the following:

Theorem 2.2. Suppose the (Hankel) matrix

$$C := \begin{bmatrix} c_0 & c_1 & \dots & c_{r-1} \\ c_1 & c_2 & \dots & c_r \\ \vdots & & & \\ c_{r-1} & \dots & & c_{2r-2} \end{bmatrix} \quad (2.2)$$

is invertible and $m = 2r - 1$. Then there are complex numbers $\{z_1, \dots, z_r\}$ and $\{a_1, \dots, a_r\}$ such that (2.1) holds. Here the z_i need not be distinct. If (say) z_1 has multiplicity 2, i.e., $z_1 = z_2$, then the terms $a_1 z_1^k + a_2 z_2^k$ must be interpreted as $(a_1 + a_2 k) z_1^k$ (and similarly for z_j of higher multiplicity).

Remark. The term "generically" means here that C in (2.2) is invertible, i.e., that (c_0, \dots, c_{2r-1}) lies off some set of zero measure, which here is a certain smooth surface in \mathbb{R}^{2r} . It is also "generically" true that the $\{z_j\}$ will all be distinct; what this requires will be clear from the proof.

Proof. Consider the system of equations

$$\begin{aligned} c_r &= b_0 c_0 + b_1 c_1 + \dots + b_{r-1} c_{r-1} \\ c_{r+1} &= b_0 c_1 + b_1 c_2 + \dots + b_{r-1} c_r \\ &\vdots \\ c_{2r-1} &= b_0 c_{r-1} + b_1 c_r + \dots + b_{r-1} c_{2r-2} \end{aligned} \quad (2.3)$$

If (2.2) holds it is solvable for b_0, \dots, b_{r-1} . That means the sequence c_0, \dots, c_{2r-1} satisfies a recurrence of length r :

$$c_n = b_{r-1} c_{n-1} + b_{r-2} c_{n-2} + \dots + b_0 c_{n-r} \quad (2.4)$$

for $n = r, r + 1, \dots, 2r - 1$ (indeed, $\{c_n\}$ can then be extended to an *infinite* sequence by (2.4), i.e., c_0, \dots, c_{2r-1} is a *section* of a sequence characterized by the recursion (2.4) ... and that is a natural way to think of what (2.1) means).

From the well-known results on solutions of difference equations it is then clear that (2.1) holds for suitable $\{a_j\}, \{z_j\}$, when the proviso about multiple $\{z_j\}$ is taken into account. However, for notational purposes let's carry out the calculation. The recurrence (2.4) has "exponential solutions" $c_n = e^{n\lambda}$ where

$$e^{n\lambda} = \sum_{k=0}^{r-1} b_k e^{(n-r+k)\lambda}$$

or

$$1 = \sum_{k=0}^{r-1} b_k \zeta^{r-k} \quad (2.5)$$

where

$$\zeta = e^{-\lambda} \quad (2.6)$$

or, more conveniently with $\omega = 1/\zeta = e^\lambda$

$$\omega^r - b_{r-1}\omega^{r-1} - b_{r-2}\omega^{r-2} - \dots - b_0 = 0. \quad (2.7)$$

Thus, the solutions of (2.4) are precisely functions of the form (2.1), where the $\{z_j\}$ are the root of (2.7), and multiple roots are interpreted as described above (henceforth we'll tacitly assume the roots are distinct).

Remark. The existence of a recurrence relation of order r implies that if the first r elements are 0 so are all remaining ones.

Note. The condition of multiple roots of the associated polynomial $z^r - b_{r-1}z^{r-1} - \dots - b_0$ is equivalent to the vanishing of a certain polynomial ("discriminant") in b_0, \dots, b_{r-1} and hence in $c_0, c_1, \dots, c_{2r-1}$ (because of (2.3)) ... which is why simple roots are "generic".

At this point, the real problems begin. Namely, the "Prony" solution even when it exists may be disadvantageous because the $\{z_j\}$ are in the "wrong" places, or the $\{a_j\}$ (in (2.1)) are too large. So we ask: can a larger (and hence "redundant") choice of r help in this respect?

Another, apparently difficult question involves "erratic sampling". Suppose we look for $\{a_j\}, \{z_j\}$, such that (2.1) holds for values of k which are *not* consecutive? This is equivalent to fitting an r -form recursive linear sequence to an "irregular" sequence of values of k . We won't try to examine this here, but rather look at the first problem.

Let's illustrate the situation with $m = 5$. We are given c_0, c_1, \dots, c_5 . Prony's solution requires solving the system (here $r = 3$):

$$\begin{aligned} c_3 &= c_0 b_0 + c_1 b_1 + c_2 b_2 \\ c_4 &= c_1 b_0 + c_2 b_1 + c_3 b_2 \\ c_5 &= c_2 b_0 + c_3 b_1 + c_4 b_2 \end{aligned}$$

for b_0, b_1, b_2 ; then solving (2.7), i.e.,

$$z^3 - b_2 z^2 - b_1 z - b_0 = 0$$

to get the $\{z_j\}$; then finding the $\{a_j\}$ by linear algebra from (2.1).

Let's now consider the "redundant" value $r = 4$. We have to satisfy

$$\begin{aligned} c_4 &= c_0 b_0 + c_1 b_1 + c_2 b_2 + c_3 b_3 \\ c_5 &= c_1 b_0 + c_2 b_1 + c_3 b_2 + c_4 b_3 \end{aligned} \quad (2.8)$$

an *underdetermined* system (in general) for b_0, b_1, b_2, b_3 . This will give us a solution with (in general) two free parameters, so that (2.7) will be a fourth-degree polynomial equation whose coefficients have two free parameters. The interesting problem is then, how do the roots of (2.7) move as these parameters are varied?

Here is a very simple *ad hoc* optimization procedure that illustrates the possibilities.

The general solution of (2.8) is

$$b_j = \beta_j + p_j s + q_j t \quad (j = 0, \dots, 3)$$

where $\vec{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)$ is any particular solution and \vec{p}, \vec{q} are 4-vectors that are a basis for solutions of the corresponding homogenous systems. The left side of (2.7) is then:

$$f(w) := w^4 - \sum_{j=0}^3 (\beta_j + p_j s + q_j t) w^j$$

or

$$f(z) = B(z) - sP(z) - tQ(z)$$

where

$$B(z) = w^4 - \sum_{j=0}^3 \beta_j w^j$$

$$P(z) = \sum_{j=0}^3 p_j w^j$$

$$Q(z) = \sum_{j=0}^3 q_j w^j$$

Thus, in this simple case the problem is: given the (monic) polynomial $B(z)$ of degree 4, and two cubic polynomials $P(z), Q(z)$, choose complex parameters such that the roots of

$$f(z) = B(z) - sP(z) - tQ(z) \quad (2.9)$$

are "where we want them", or close to this set, e.g., in \mathbb{D} , near \mathbb{D} , etc. A very simple starting point is to notice that

$$\phi(s, t) := \frac{1}{2\pi} \int_0^{2\pi} \log |f(e^{i\theta})| d\theta$$

is $\sum \log |z_j|$, summed over the roots $\{z_j\}$ of f lying in $\{|z| > 1\}$.

It is an easy problem to minimize (numerically) $\phi(s, t)$. If there is a solution with all roots in \mathbb{D} this procedure will find one.

Similar procedures can of course be applied with larger m and r .

We might also want the $\{z_j\}$ to be *on* the unit circle (or near it). Even in the above simple-minded example that seems to lead to nontrivial optimization problems. The general theme seems to be *controlling the roots of polynomials whose coefficients contain parameters*.

There's lots of room to experiment here. For example, instead of minimizing the geometric mean of the polynomial $f(z) = f(z; s, t)$ given by (2.9) we could minimize $\|f(\cdot; s, t)\|_p$ for some small positive p . This is a subject for continuing research.

3. References

- [BM75] M. Van Blaricum and R. Mittra. A technique for extracting the poles and residues of a system directly from its transient response. *IEEE Trans. Antennas Propagation*, 23:777-781, 1975.
- [HS89] Y. Hua and T.K. Sarkar. A discussion of E-pulse method and Prony's method for radar target resonance retrieval from scattered field. *IEEE Trans. Antennas Propagation*, 37(7):944-946, 1989.
- [JSW83] V.K. Jain, T.K. Sarkar, and D.D. Weiner. Rational modeling by the pencil-of-function method. *IEEE Trans. Acoustics, Speech, Signal Processing*, 31:564-573, 1983.
- [KM65] E.M. Kennaugh and D.L. Moffat. Transient and impulse response approximations. *Proc. IEEE*, 53:893-901, 1965.
- [Per75] A.J. Perni. Target identification by natural resonant estimation. *IEEE Trans. Aerospace Electronic Systems*, 11(2):147-154, 1975.
- [RCN87] E.J. Rothwell, K.M. Chen, and D.P. Nyquist. Extraction of the natural frequencies of a radar target from a measured response using E-pulse techniques. *IEEE Trans. Antennas Propagation*, 35(6):715-720, 1987.
- [S⁺80] T.K. Sarkar et al. Suboptimal approximation/identification of transient waveform from electromagnetic systems by pencil-of-functions method. *IEEE Trans. Antennas Propagation*, 28(6):928-932, 1980.

Prometheus Inc.
Final Report
26 November 1991

Section IX

Barker sequences and Littlewood's "two-sided conjectures" on polynomials with ± 1 coefficients

BARKER SEQUENCES AND LITTLEWOOD'S "TWO-SIDED CONJECTURES" ON POLYNOMIALS WITH ± 1 COEFFICIENTS*

B. SAFFARI†
J.S. BYRNES‡
Prometheus Inc.
21 Arnold Ave.
Newport, RI 02840

1. Introduction

One purpose of this note is to mention two conjectures going back to the 1960's, both still unproved to the end of 1990, and to give a very simple proof of the (apparently unnoticed) fact that *at least one of them is true*. These two conjectures are :

(A) TURYN'S CONJECTURE (around 1960 ?) : There are only a finite number of binary Barker sequences. (The definition of Barker sequences will be recalled in § 3).

(B) LITTLEWOOD'S TWO-SIDED CONJECTURE (1966) : For infinitely many integers n , there is a polynomial P of degree n , with *all* its coefficients equal to ± 1 , such that

$$A\sqrt{n+1} \leq |P(e^{it})| \leq B\sqrt{n+1} \quad (\text{on the whole unit circle})$$

where A and B are positive *absolute* constants.

This note may also be considered as the embryo of a somewhat expository note, in that we first recall (in § 3) a few facts about Barker sequences ; then, in § 4, we recall the two versions (i.e., strong and weak) of Littlewood's two-sided conjectures ; in § 5 we mention some partial results for the Littlewood conjectures ; in § 6 we prove the theorem mentioned at the beginning of this introduction.

We wish to thank J.-L. Nicolas and A. M. Odlyzko for bibliographical information.

2. Terminology

We define the *length* $L(P)$ of a polynomial $P(z) = \sum_{k=0}^n a_k z^k$ with complex coefficients by

$$L(P) = \sum_{k=0}^n |a_k|.$$

* This work was supported by AFOSR, under contract F-49620-90-C-0023.

† Also at Université de Paris-Sud, Dept. de Mathématiques.

‡ Also at the University of Massachusetts at Boston.

This follows the standard terminology (of Number Theory). We also set, as usual,

$$\|P\|_2 = \left(\frac{1}{2\pi} \int_0^{2\pi} |P(e^{it})|^2 dt \right)^{1/2} = \left(\sum_{k=0}^n |a_k|^2 \right)^{1/2},$$

and

$$\|P\|_\infty = \max_{0 \leq t \leq 2\pi} |P(e^{it})|.$$

P is called a *unimodular polynomial* if all its coefficients a_k have modulus one. Thus, for any unimodular polynomial,

$$\begin{aligned} L(P) = \|P\|_2^2 &= \deg P + 1 = \text{number of terms in } P(z) \\ &= \text{length of the finite sequence } \{a_0, a_1, \dots, a_n\}. \end{aligned}$$

P is called a ± 1 polynomial if $a_k = \pm 1$ for all $k = 0, 1, \dots, n$.

3. Barker sequences

Historically, such sequences arose from telecommunication engineering. Shortly after the publication of Barker's paper [1] in 1953, it became customary to say that a finite sequence $\{a_0, a_1, \dots, a_n\}$ all of whose terms are ± 1 is a *Barker sequence* if the autocorrelation coefficients

$$(1) \quad c_j = \sum_{k=0}^{n-j} a_k a_{k+j}$$

satisfy

$$|c_j| \leq 1 \text{ for all } j = 1, 2, \dots, n$$

or, equivalently,

$$c_j = \pm 1 \text{ or } 0 \quad (\text{for } j = 1, 2, \dots, n)$$

since here c_j is an integer for all j . (The "central" autocorrelation c_0 satisfies, of course, $c_0 = n + 1 = \text{length of the } \pm 1 \text{ sequence}$).

It has been proved by Storer and Turyn [21], and independently by Poliak and Moshetov [16], that there are only five possible odd values ≥ 3 for the length $L = n + 1$, namely $L = 3, 5, 7, 11, 13$. As for even lengths, there are only two such values of L which have been discovered so far, namely $L = 2$ and $L = 4$. Turyn has shown that if there is another Barker sequence (necessarily of even length), then L must be quite large. Turyn's work has been somewhat refined by recent (unpublished) work. There is also overwhelming theoretical evidence (among which D. J. Newman's conjecture for the L^1 norm of ± 1 polynomials, which we hope to treat in detail elsewhere) for the truth of the following conjecture:

CONJECTURE (Turyn and others) : *There are only a finite number of Barker sequences.*

This conjecture is generally believed to be true, and also its stronger form which asserts that *there are no Barker sequences of even length > 4 .*

The Barker sequences (and their generalizations) have been the object of much research and many publications, both from the communication engineering view-point and the pure mathematics view-point. We shall not attempt here to give a list of those contributions. Instead, we just mention a couple of open problems (conjectures) on "generalized Barker sequences". A finite sequence $\{a_0, a_1, \dots, a_n\}$ of complex numbers is called a "generalized Barker sequence" if $|a_k| = 1$ for all $k = 0, 1, \dots, n$ and if the autocorrelation coefficients

$$c_j = \sum_{k=0}^{n-j} \bar{a}_k \cdot a_{k+j}$$

satisfy

$$|c_j| \leq 1 \text{ for all } j = 1, 2, \dots, n.$$

(Again, $c_0 = n + 1$ obviously). A generalized Barker sequence all of whose terms are d^{th} roots of 1 (for a fixed integer $d \geq 2$) is called a "*d-phase Barker sequence*". In the case $d = 2$, the "classical" ± 1 Barker sequences mentioned at the beginning of this § 3 are naturally referred to as "binary Barker sequences". (In general, a binary sequence is any sequence taking only two values).

To the best of our knowledge, for no value of the integer $d \geq 2$ it is known whether there are infinitely many or only a finite number of *d-phase Barker sequences*. We conjecture that the number of *d-phase Barker sequences* is finite (but, of course, depends on *d*).

On the other hand one would conjecture that the set of lengths of all possible generalized Barker sequences is unbounded *if no additional condition whatsoever is imposed*. An additional condition may change things drastically. Thus we have the following result :

THEOREM. *If a generalized Barker sequence $\{a_0, a_1, \dots, a_n\}$ is self-inversive, that is, satisfies*

$$(2) \quad a_{n-k} = \bar{a}_k \text{ for all } k = 0, 1, \dots, n$$

then the length $n + 1$ cannot exceed 2, (that is, we have $n = 0$ or $n = 1$).

This result and some of its generalizations are corollaries of a 1989 theorem due to Fredman, Saffari and Smith [8]. (The paper [8] is in French, because of its appearing in the Comptes Rendus de l'Académie des Sciences de Paris, but the first author also wrote an English translation of the manuscript). Yet, in contrast to the above result on *self-inverse* generalized Barker sequences, the following problem remains open :

PROBLEM. Do there exist generalized Barker sequences $\{a_0, a_1, \dots, a_n\}$ of arbitrarily large lengths which are *palindromic* (or *symmetric*), i. e. such that

$$(3) \quad a_{n-k} = a_k \quad (\text{for all } k = 0, 1, \dots, n) ?$$

This problem was mentioned to the first author in 1988 by Dr. Ning Zhang and is due (if his memory is faithful) to Prof. S. W. Golomb (or to S. W. Golomb and R. A. Scholtz). Anyway, despite our pledge not to give in this note any references to work on Barker sequences, we cannot refrain from quoting the very interesting 1965 paper [11] of Golomb and Scholtz.

Let us emphasize that conditions (2) and (3) (i.e. self-inversiveness and symmetry) are, in fact, very different and require different methods of attack.

To finish this § 3, let us return to the *binary* Barker sequences and just mention a few additional facts about such sequences :

FACT 1. If there is any (binary) Barker sequence of length $L > 13$, then L is of the form

$$(4) \quad L = 4s^2 \quad \text{where } s \text{ is some odd integer.}$$

This has been proved in 1965 by Turyn [22] on considering (v, k, λ) cyclic difference sets related to Barker sequences.

FACT 2. The length L of a (binary) Barker sequence *cannot* be of the form

$$L = 4p^\alpha \quad \text{where } p \text{ is a prime number.}$$

This is due to Richard Turyn, but we cannot find the reference (and unfortunately we did not contact Turyn to ask him). This reference will be given, however, in a forthcoming expository paper [14] by S. Y. R. Li and Ning Zhang.

FACT 3. If there is any (binary) Barker sequence of length $L > 13$, then L is not divisible by any prime $p \equiv 3 \pmod{4}$.

This is proved, as a corollary of a much more general result on "Golay complementary sequences", in the 1990 paper [5] by Eliahou, Kervaire and Saffari. A much simpler proof of this same result (again, in the more general context of Golay sequences) will appear in a forthcoming paper [6] of Eliahou, Kervaire and Saffari. The first author presented an

oral exposition of that very nice simple proof of Kervaire in the Harmonic analysis seminar at Orsay [19]. The previous proof was due to Shalom Eliahou, who improved on weaker results of Kervaire and the first author.

The next "Fact" assumes the definition of the monoid of Golay pairs of complementary sequences. For a quite thorough treatment of the subject we refer to [6]. Since "Fact 4" below is unrelated to the analytic problems of § 4, 5 and 6, the reader may ignore "Fact 4" and move to § 4.

FACT 4. If there is any binary Barker sequence of length $L > 13$ (so that $L = 4s^2$, s odd, by Fact 1), then there exists a Golay pair (G_1, G_2) of common length $L/2 = 2s^2$, and (G_1, G_2) is necessarily irreducible in the monoid of all Golay pairs.

COROLLARY (to Facts 1, 2, 4) :

(A) *If the monoid of all Golay pairs is finitely generated, then there are only a finite number of binary Barker sequences. In fact, if the monoid of all Golay pairs contains only a finite number of (necessarily irreducible) elements whose length is of the form $2s^2$, where $s > 3$ is odd and not a prime power, then there are only a finite number of binary Barker sequences.*

(B) *If the monoid of all Golay pairs has no (necessarily irreducible) element whose length is of the form $2s^2$, with $s \geq 65$, odd and not a prime power, then there are no binary Barker sequences of length > 13 . (Remark : $s \geq 65$ implies $2s^2 \geq 8450$).*

REMARK. The argument leading to part (B) of this corollary also yields : *If there is any binary sequence of length $L > 13$, then necessarily $L \geq 16900$. In fact, the sequence of candidates for L reads : {16 900, 28 900, 84 100, 136 900, 168 100, etc.}.*

4. Littlewood's "two-sided conjectures"

The weak version : For infinitely many integers n , there is some ± 1 polynomial P of degree n , such that

$$A\sqrt{n+1} \leq |P(e^{it})| \leq B\sqrt{n+1} \quad (\text{on the whole unit circle})$$

where A and B are positive absolute constants.

The strong version : For every integer $n \geq 2$, there is some ± 1 polynomial P of degree n , such that

$$C\sqrt{n+1} \leq |P(e^{it})| \leq D\sqrt{n+1} \quad (\text{on the whole unit circle})$$

where C and D are positive absolute constants.

Both of these versions are due to Littlewood [15]. There are several reasons for considering the weak version as well as the strong version, among which :

(A) Even the “weak version” remains unsolved to this date (end of 1990).

(B) Quite conceivably the “weak version” might hold true with *better* constants (A and B) than the respective constants (C and D) of the “strong version”. One reason for suspecting such a phenomenon is the analogy with the supremum norm conjectures for ± 1 polynomials on the unit circle. (Cf. the appendix). Thus, if in addition to a proof of Littlewood’s two-sided conjecture(s) we also look for best possible constants, then the distinction between the weak and strong versions might be essential.

(C) Assuming the truth of the weak conjecture, we do obtain some partial information for every $n \geq 2$, as shown in our Theorem 1 (in § 5).

5. Partial results for Littlewood’s “two-sided conjectures”

There are several such results which deserve to be recalled in a systematic exposition of the subject. In this note we just quote three breakthroughs, due to Körner [13], Kahane [12] and Beck [2].

KÖRNER’S THEOREM. *For all integer $n \geq 2$, there is a unimodular polynomial P of degree n such that :*

$$C\sqrt{n+1} \leq |P(e^{it})| \leq D\sqrt{n+1} \quad (\text{on the whole unit circle})$$

where C and D are absolute positive constants.

This theorem of Körner actually proved another conjecture of Littlewood formulated in the same paper [15], which was the analogue of (the strong version of) the two-sided conjecture, but for the class of unimodular polynomials instead of the class of ± 1 polynomials. For the latter class the problem appears to be much tougher. Körner used a construction of Byrnes [3] and probabilistic ideas. That same year (1979-1980) Kahane [12] substantially refined Körner’s method and proved the following extremely surprising result :

KAHANE’S THEOREM. *There is a sequence $\epsilon_n > 0$ with $\lim_{n \rightarrow \infty} \epsilon_n = 0$, such that for each integer $n \geq 2$ there is some unimodular polynomial P satisfying*

$$(1 - \epsilon_n)\sqrt{n+1} \leq |P(e^{it})| \leq (1 + \epsilon_n)\sqrt{n+1} \quad (\text{on the whole unit circle}).$$

Actually Kahane obtained *quantitative* estimates for ϵ_n , both in [12] and afterwards.

A by-product of Kahane's theorem is the disproof of the 1957 conjecture of Erdős on the sup-norm of unimodular polynomials (see Problem 22 in [7]). The Erdős conjecture and some of its variations will be dealt with elsewhere. (See also [8]). The first author coined the term "Kahane's ultra-flat polynomials" to describe the unimodular polynomials P of Kahane's theorem. They have some remarkable properties and will be studied elsewhere.

Very recently (May 1990) Beck added more sophisticated ingredients to Kahane's method and obtained [2] :

BECK'S THEOREM. *For sufficiently large (integer) d , say $d \geq 300$, and sufficiently large $n \geq n_0(d)$, there is a polynomial P of degree n , all of whose coefficients are d^{th} roots of 1, such that :*

$$C\sqrt{n+1} \leq |P(e^{it})| \leq D\sqrt{n+1} \quad (\text{on the whole unit circle})$$

where C and D are absolute positive constants.

The value $d = 300$ in Beck's theorem can be lowered somewhat, but the theoretical limit to his approach is $d = 3$, so the case $d = 2$ (that is, Littlewood's two-sided conjecture for ± 1 polynomials) probably requires completely new ideas.

It is worth noting that all of these remarkable three theorems (of Körner, Kahane and Beck) are "ineffective", in that they prove the existence of P by probabilistic methods without providing explicit constructions of such P .

For ± 1 polynomials there are several partial results in the literature, but those pertaining to the lower bound are quite weak (except in [4], but there simultaneous upper bound conditions are not considered). Here is an observation (of the first author, unless it is not new) establishing a (very loose) link between the two versions (i.e. weak, resp. strong) of Littlewood's two-sided conjecture :

THEOREM 1. *Suppose that the weak form of Littlewood's two-sided conjecture for ± 1 polynomials is true. Then there is a sequence $\epsilon_n > 0$ with $\lim_{n \rightarrow \infty} \epsilon_n = 0$ such that for all integers $n \geq 2$ we can find some ± 1 polynomial of degree n satisfying :*

$$(n+1)^{1/2-\epsilon_n} \leq |P(e^{it})| \leq (n+1)^{1/2+\epsilon_n} \quad (\text{on the whole unit circle}).$$

This result is extremely unsatisfactory as far as the upper bound is concerned). (If the lower bound condition is ignored, one has of course the upper bound $B\sqrt{n+1}$, B = some absolute constant, see the appendix). For this reason we shall not prove Theorem 1 here, but hope to replace it by a sharper result in a new version of this note. Let us nevertheless mention that the proof of (the present form of) Theorem 1 uses largely the method of proof of Theorems 2.4 and 2.5 of the interesting paper [4].

6. Barker polynomials of length greater than 13 : An analytic observation.

To any Barker sequence $\{a_0, a_1, \dots, a_n\}$ we assign the "Barker polynomial":

$$(5) \quad P(z) = \sum_{k=0}^n a_k z^k.$$

We now state and prove the result mentioned at the beginning of the introduction :

THEOREM 2. *Either there are only a finite number of binary Barker sequences. Or else, if there is an infinite sequence of binary Barker sequences of respective lengths $L_1, L_2, \dots, L_p, \dots$, with $\lim_{p \rightarrow \infty} L_p = \infty$, then the corresponding Barker polynomials given by (5) (with of course $n+1 = L_p$) satisfy:*

$$(6) \quad A_n \sqrt{n+1} \leq |P(e^{it})| \leq B_n \sqrt{n+1} \quad (\text{on the whole unit circle})$$

with

$$(7) \quad \begin{cases} \lim_{n \rightarrow \infty} A_n = \sqrt{1-\theta} = 0.579... \\ \lim_{n \rightarrow \infty} B_n = \sqrt{1+\theta} = 1.289... \end{cases}$$

where

$$(8) \quad \theta = \sup_{\alpha > 0} \frac{(1 - \cos \alpha) \sin \alpha}{\alpha} = \sup_{\beta > 0} \frac{2 \cos \beta \sin^3 \beta}{\beta} = 0.6639...$$

(6) implies an affirmative answer to the weak form of Littlewood's problem, with constants A and B as close as we wish to the two respective numerical values (7).

Proof of Theorem 2. It is extremely simple. Let P be our Barker polynomial given by (5), with n arbitrarily large. Of course n is odd, since the length $L = n+1$ must be even. Let us recall two facts about P . The first fact is that (for $n \geq 3$) L is a multiple of 4 :

$$(9) \quad L = 4m.$$

Of course (9) is just a weak form of (4), but for the reader's convenience we shall give here a straightforward (and well known) direct proof of (9). It is as follows : from the definition of Barker sequences we immediately see that the autocorrelation coefficients c_j satisfy, for even L (i.e., odd n) :

$$(10) \quad \begin{cases} c_j = 0 & \text{for even } j \neq 0 \\ c_j = \pm 1 & \text{for odd } j \quad (j = 1, 3, \dots, n). \end{cases}$$

We only use the top line equalities in (10) (i.e., $c_j = 0$ for even $j \neq 0$). On setting $j = 2\ell$ and on splitting the sum in the right side of (1) as two sums corresponding to even and odd subscripts k , we can rewrite the top line equalities in (10) as :

$$(11) \quad \sum_{h=0}^{(n-1-2\ell)/2} a_{2h} a_{2h+2\ell} + \sum_{h=0}^{(n-1-2\ell)/2} a_{2h+1} a_{2h+1+2\ell} = 0 \quad (\text{for all } \ell = 0, 1, \dots, \frac{n-1}{2}).$$

(11) means that the two finite ± 1 sequences $\{a_0, a_2, \dots, a_{n-1}\}$ and $\{a_1, a_3, \dots, a_n\}$ of common length $(n+1)/2$ form a pair of *Golay complementary sequences*. Equivalently, if we write $P(z)$ as a sum of an even polynomial and an odd polynomial :

$$(12) \quad P(z) = S(z^2) + zT(z^2),$$

then the fact that $\{a_0, a_2, \dots, a_{n-1}\}$ and $\{a_1, a_3, \dots, a_n\}$ form a Golay pair of complementary sequences is equivalent to the identity

$$(13) \quad |S(\zeta)|^2 + |T(\zeta)|^2 = n+1 \quad (\text{for all } \zeta \in \mathbb{C} \text{ with } |\zeta| = 1)$$

and also to the identity

$$(14) \quad |P(\zeta)|^2 + |P(-\zeta)|^2 = 2(n+1) \quad (\text{for all } \zeta \in \mathbb{C} \text{ with } |\zeta| = 1).$$

Now it is well known (and easily proved, by Golay himself [10] as early as 1951) that the common length of a pair of Golay complementary sequences is even. Thus $(n+1)/2$ is even, hence $n+1 \equiv 0 \pmod{4}$, hence the desired relation (9).

The second fact we use (Storer and Turyn [21]) is this : with $n = 4m$, we have

$$(15) \quad c_{4m-j} = -c_j \quad (\text{for all } j = 1, 2, \dots, 4m-1).$$

Of course c_j is of interest for odd values of j only, since here $c_j = 0$ for even $j \neq 0$. (15) easily follows from the definition of the autocorrelations c_j and by using the following straightforward lemma : If $\{\epsilon_1, \epsilon_2, \dots, \epsilon_N\}$ is a ± 1 sequence of length N , then

$$(16) \quad \epsilon_1 \epsilon_2 \dots \epsilon_N = (-1)^{(N-S)/2} \quad \text{where } S = \epsilon_1 + \epsilon_2 + \dots + \epsilon_N.$$

Let us now prove Theorem 2. We have, for all real t ,

$$(17) \quad |P(e^{it})|^2 = \sum_{k=-n}^n c_k e^{ikt}.$$

Here the central autocorrelation coefficient is $c_0 = n+1 = L = 4m$, and otherwise $c_{-k} = c_k$. Thus, on setting

$$(18) \quad f(t) = |P(e^{it})|^2 - 4m,$$

we have from (17) :

$$\begin{aligned}
f(t) &= 2 \sum_{k=1}^n c_k \cos kt \quad (\text{with } n = 4m - 1) \\
&= 2 \sum_{k=1}^{2m-1} c_k (\cos kt - \cos(4m - k)t) \quad (\text{because of (15)}) \\
&= 2 \sum_{h=1}^m c_{2h-1} (\cos(2h-1)t - \cos(4m - 2h + 1)t) \\
&\quad (\text{because } c_j = 0 \text{ for even } j \neq 0) \\
&= 4 \sum_{h=1}^m c_{2h-1} \sin 2mt \sin(2m - 2h + 1)t \\
&= 4(\sin 2mt) \sum_{h=1}^m c_{2m-2h+1} \sin(2h-1)t.
\end{aligned}$$

We use this to crudely majorize $|f(t)|$:

$$|f(t)| \leq 4 |\sin 2mt| \cdot \sum_{h=1}^m |\sin(2h-1)t| \leq 4m\theta_m$$

where

$$\theta_m = \max_{0 \leq t \leq 2\pi} \left(\frac{1}{m} |\sin 2mt| \cdot \sum_{h=1}^m |\sin(2h-1)t| \right).$$

A computation shows that

$$\begin{aligned}
\lim_{m \rightarrow \infty} \theta_m &= \sup_{\alpha > 0} \left(|\sin \alpha| \cdot \int_0^1 |\sin \alpha x| dx \right) \\
&= \sup_{\gamma > 0} \frac{(1 - \cos \gamma) \sin \gamma}{\gamma} \\
&= \sup_{\beta > 0} \frac{2 \cos \beta \sin^3 \beta}{\beta} = 0.6639 \dots
\end{aligned}$$

whence (6) with $A_n = \sqrt{1 - \theta_m}$, $B_n = \sqrt{1 + \theta_m}$.

This completes the proof of Theorem 2.

REMARK. By refining the above calculations one can replace the numerical constants of equalities (7) and (8) by better constants.

7. Appendix : Upper bound problems for ± 1 polynomials on the unit circle.

Let

$$K_n = \min_P (\|P\|_\infty / \|P\|_2) = \min_P (\|P\|_\infty / \sqrt{n+1})$$

where the minimum is taken over the set of all ± 1 polynomials P of given degree n . There are many interesting (and mostly open) problems on the behaviour of K_n , among which :
What are the exact values of

$$K^+ = \limsup_{n \rightarrow \infty} K_n \quad \text{and} \quad K^- = \liminf_{n \rightarrow \infty} K_n ?$$

(Trivially $1 \leq K^- \leq K^+$). Since the exact determination of K^+ and K^- seems exceedingly difficult, as sharp as possible upper bounds and lower bounds for *both* of K^+ and K^- would be most welcome.

Lower bound results are so far quite poor : The *Erdős conjecture* [7] of 1957 for ± 1 polynomials, that is,

$$(19) \quad K^- > 1$$

is still unsolved, and so is the weaker conjecture

$$(20) \quad K^+ > 1.$$

The *upper bound* results are less disappointing. Shapiro [20] proved in 1951 (via the consideration of the famous "Rudin-Shapiro polynomials", known to Golay [9], [10] in a different form as early as 1949, and rediscovered by Rudin [17] in 1959) that

$$(21) \quad K^- \leq \sqrt{2} \quad \text{and} \quad K^+ \leq 2 + \sqrt{2}.$$

The first author proved [18] in 1987 that

$$K^+ \leq \sqrt{2},$$

but he no longer believes his conjecture that $K^+ = \sqrt{2}$. There is some evidence that $K^+ < \sqrt{2}$.

It is *not* known whether $K^+ = K^-$. (If we had $K^+ \neq K^-$, the truth of Conjecture (20) would of course follow !). No upper bound better than (21) is known for K^- . If there were infinitely many Barker sequences (which is very unlikely), then by Theorem 2 of this note we would have $K^- \leq 1.289 \dots$.

These problems on K_n and related matters on ± 1 polynomials will be treated in detail elsewhere.

References

- [1] R. H. BARKER *Group synchronizing of binary digital systems*, in "Communication Theory", pp. 273-287. London, Butterworth, 1953.
- [2] J. BECK "Flat" polynomials on the unit circle - on a problem of Littlewood. To appear.
- [3] J. S. BYRNES *On polynomials with coefficients of modulus one*. Bull. London Math. Soc. 9 (1977), 171-176.
- [4] F. W. CAROLL, D. EUSTICE & T. FIGIEL *The minimum modulus of polynomials with coefficients of modulus one*. J. London Math. Soc. 16 (1977), 76-82.
- [5] S. ELIAHOU, M. KERVAIRE & B. SAFFARI *A new restriction on the lengths of Golay complementary sequences*. J. Combinatorial Theory, series A, 55 (1990), 49-59.
- [6] S. ELIAHOU, M. KERVAIRE & B. SAFFARI *On Golay polynomial pairs*. Advances Appl. Math. (to appear, probably in 1991).
- [7] P. ERDOS *Some unsolved problems*. Michigan Math. J. 4 (1957), 291-300.
- [8] M. L. FREDMAN, B. SAFFARI & B. SMITH *Polynômes réciproques : conjecture d'Erdős en norme L^4 , taille des autocorrélations et inexistence des suites de Barker*. C. R. Acad. Sc. Paris 308 (1989), 461-464.
- [9] M. J. E. GOLAY *Multislit spectrometry*. J. Opt. Soc. America 39 (1949), 437-444.
- [10] M. J. E. GOLAY *Static multislit spectrometry and its application to the panoramic display of infrared spectra*. J. Opt. Soc. America 41 (1951), 468-472.
- [11] S. W. GOLOMB & R. A. SCHOLTZ *Generalized Barker sequences*. IEEE Trans. Information Theory, IT-11 (1965), 533-537.
- [12] J.-P. KAHANE *Sur les polynômes à coefficients unimodulaires*. Bull. London Math. Soc. 12 (1980), 321-342.
- [13] T. KORNER *On a polynomial of J. S. Byrnes*. Bull. London Math. Soc. 12 (1980), 219-224.

- [14] S. Y. R. LI & NING ZHANG *Expository paper on Barker sequences* (in preparation).
- [15] J. E. LITTLEWOOD *On polynomials $\sum \pm z^m$, $\sum \exp(\alpha_m i) z^m$, $z = e^{\theta i}$* . J. London Math. Soc. 41 (1966), 367-376.
- [16] Iu. V. POLIAK & R. V. MOSHETOV *Scientific transactions of the Radio-Engineering Institute of the Acad. Sci. USSR* 1, 124-159.
- [17] W. RUDIN *Some theorems on Fourier coefficients*. Proc. Amer. Math. Soc. 10 (1959), 855-859.
- [18] B. SAFFARI *Structure algébrique sur les couples de Rudin-Shapiro. Problème extrémal de Salem sur les polynômes à coefficients ± 1* . C. R. Acad. Sc. Paris 304 (1987), 127-130.
- [19] B. SAFFARI *Analytic problems related to Golay sequences and Rudin-Shapiro polynomials*. Sémin. Anal. Harm. Orsay, 1989.
- [20] H. S. SHAPIRO *Extremal problems for polynomials and power series*. M. S. Thesis, MIT, 1951.
- [21] J. STORER & R. TURYN *On binary sequences*. Proc. Amer. Math. Soc. 12 (1961), 394-399.
- [22] R. TURYN *Character sums and difference sets*. Pacific J. Math. 15 (1965), 319-346.

Prometheus Inc.
Final Report
26 November 1991

Section X
A note on rational approximations to the Fresnel integral

A note on rational approximations to the Fresnel integral†

W.H.J. Fuchs

Prometheus Inc.
21 Arnold Avenue
Newport, RI 02840

Also at Cornell University.

W.K. Hayman

Prometheus Inc.

Also at University of York.

J.S. Byrnes

Prometheus Inc.

Also at University of Massachusetts at Boston.

1. Introduction

In some problems of applied mathematics, the evaluation of the Fresnel integral

$$f(\omega) = e^{i\omega^2} \int_{\omega}^{\infty} e^{-it^2} dt$$

is needed. It is therefore of interest to have easily computable approximations to $f(\omega)$, primarily for positive values of ω , but also for complex ω .

The function $f(\omega)$ has been studied by many authors and in particular by J.B. Rosser in the monograph [Ros48].

In diffraction analysis, a closed-form estimate for the modified Fresnel integral,

$$K_{\pm}(\chi) = \frac{1}{\sqrt{\pi}} F_{\pm}(\chi) e^{\mp i(\chi^2 + \pi/4)}, \quad \text{where}$$

$$F_{\pm}(\chi) = \int_{\chi}^{\infty} e^{\pm it^2} dt,$$

is often required. The error in the James approximation [Jam79]

$$K_{\pm}(\chi) \approx \frac{1}{2\sqrt{\pi\chi^2 + \chi + 1}} e^{\pm i(\tan^{-1}(\chi^2 + 1.5\chi + 1) - \pi/4)}, \quad \text{for } \chi \geq 0$$

is as much as two percent for the phase component and eight percent for the amplitude component. Abuelma'atti [Abu89] gives numerical evidence, but no proof, indicating that his approximation,

$$K_{\pm}(\chi) \approx \left[\frac{1}{2\sqrt{\pi\chi^2 + \chi + 1} + \alpha\chi e^{-\beta\chi^2}} \right] e^{\pm i(\tan^{-1}(\chi^2 + 1.5\chi + 1) - \pi/4)}, \quad (\chi \geq 0),$$

† This research was supported by AFOSR under contract F49620-90-C-0023.

where $\alpha = 0.022$ and $\beta = 0.29$, improves the result of James by yielding an error which is at most two percent for the phase component and 3.4 percent for the amplitude component. We offer a different approach.

In section 2, we investigate the approximation of $f(\omega)$ by rational fractions. We also present numerical results for approximation by rational fractions of low degree.

2. Results

By [Ros48, (1.17), p. 7; Theorem 1.2, p. 12],

$$\sqrt{\pi}f(\omega) = \int_0^\infty \frac{e^{-\omega^2 t^2}}{t^2 + i} dt \quad (|\arg \omega| < \pi/4) \quad (2.1)$$

If $R(s)$ is a rational function satisfying

$$|e^{-s} - R(s)| < \epsilon \quad (0 \leq s), \quad (2.2)$$

then, for $\omega > 0$,

$$\begin{aligned} \left| \int_0^\infty (e^{-\omega^2 t^2} - R(\omega^2 t^2))(t^2 + i)^{-1} dt \right| &< \epsilon \int_0^\infty |t^2 + i|^{-1} dt \\ &\leq \epsilon \int_0^\infty (1 + t^4)^{-1/2} dt \\ &= \frac{\epsilon}{4\sqrt{\pi}} \cdot \Gamma^2\left(\frac{1}{4}\right). \end{aligned} \quad (2.3)$$

By (2.2), $R(s)$ is of the form

$$R(s) = a_0 + \sum_{k=1}^m \sum_{l=1}^{n_k} a_{kl}(s - \sigma_k^2)^{-l},$$

σ_k in the upper half-plane; therefore,

$$r(\omega) = \frac{1}{\sqrt{\pi}} \int_0^\infty R(\omega^2 t^2)(t^2 + i)^{-1} dt$$

is equal to a finite linear combination of terms of the form

$$I_l = \int_0^\infty (\omega^2 t^2 - \sigma^2)^{-l} (t^2 + i)^{-1} dt.$$

By an easy application of the residue theorem,

$$I_1 = 1/2 \int_{-\infty}^\infty (\omega^2 t^2 - \sigma^2)^{-1} (t^2 + i)^{-1} dt = \frac{\pi}{2\sigma} \frac{1}{\omega - \eta\sigma} \quad (\eta = e^{ni/4}).$$

I_l is obtained from I_1 by applying the operator $1/(l-1)! (\frac{1}{2\sigma} \frac{d}{d\sigma})^{l-1}$ and is therefore a rational function with a pole of order l at $\eta\sigma$ and with no other poles. It follows that $r(\omega)$ is a rational function of order n with poles at the points $\eta\sigma_k$. By (2.2) and (2.3),

$$|f(\omega) - r(\omega)| < \epsilon \Gamma^2(1/4)/(4\pi). \quad (2.4)$$

We now quote a known result ([GR39]; see also [Var86]):

Lemma 2.1. There are rational functions $R_n(z)$ of degree n such that

$$\left(\sup_{s \geq 0} |e^{-s} - R_n(s)|\right)^{1/n} \rightarrow \lambda = 1/9.289 \dots$$

Letting $R(s) = R_n(s)$, we obtain

Theorem 2.2. For $n > n_0(\delta)$ one can find a rational function $r_n(\omega)$ such that

$$\sup_{\omega \geq 0} |f(\omega) - r_n(\omega)| < (\Gamma^2(1/4)/4\pi)(\lambda + \delta)^n \quad \text{for } \delta > 0. \quad (2.5)$$

We conjecture that the $\lambda + \delta$ on the right hand side of (2.5) cannot be replaced by any number $< \lambda$.

We have also considered other ways of choosing $R(s)$ in (2.2). In particular, for

$$R(s) = (1 + s/n)^{-n}$$

we found:

Theorem 2.3. If

$$r_n(\omega) = \sqrt{\pi} \eta 4^{-n} \sum_{k=0}^{n-1} \binom{n+k-1}{k} X^{k-n},$$

$$X = (1 + (\eta\omega/n^{1/2}))/2,$$

then, in

$$-(3\pi/4) + \delta \leq \arg \omega \leq (\pi/4) - \delta \quad \text{for } 0 < \delta < \pi,$$

$$|f(\omega) - r_n(\omega)| < .6/(n \max(|\omega|, 1) \sin \delta) \quad (n \geq 4)$$

and the .6 can be replaced by .252, if $n \geq 100$.

The explicit formula for $r_n(\omega)$ is achieved here at the expense of the much larger error term.

3. References

- [Abu89] Muhammad Taher Abuelma'atti. An improved approximation to the Fresnel integral. *IEEE Trans. Antennas Propagation*, 37(6):946-947, 1989.
- [GR39] A.A. Gonchar and E.A. Rakhmanov. Distribution and the rate of rational approximation of analytic functions. *USSR-Sb62*, 2:305-348, 1939.
- [Jam79] G.L. James. An approximation to the Fresnel integral. *Proc. IEEE*, 67:677-678, 1979.
- [Ros48] J.B. Rosser. *Theory and Application of $e^{iz^2} \int e^{-iz^2} dx$* . Mapleton House, 1948. Part I.
- [Var86] R.S. Varga. Scientific computation on some mathematical conjectures. In C.H. Chin, L.L. Shumaker, and J.D. Ward, editors, *Approximation Theory V*. Academic Press, 1986.

Prometheus Inc.
Final Report
26 November 1991

Section XI
Personnel and papers presented

Applications of Approximation Theory in Antenna Design, Signal Processing and Filtering

Personnel

The Prometheus personnel who performed the research reported herein were:

Dr. 's James S. Byrnes, W. H. J. Fuchs, W. K. Hayman, T. W. Körner, Donald J. Newman, Gerald Ostheimer, Richard Roy, B. Saffari, and H. S. Shapiro.

Presentations

During the period of performance on this contract, the Principal Investigator Dr. James S. Byrnes gave presentations regarding this work at:

Lawrence Livermore National Laboratory, Naval Research Laboratory, BDM, MRJ Inc., Defense Sciences Technical Organization (Adelaide, Australia), Commonwealth Science and Industrial Research Organization (Sydney, Australia), an international NATO Advanced Study Institute in Pisa, an international signal processing conference at Cetraro, Italy, the Universities of Montreal, Adelaide, New South Wales, Paris, Rome, Pisa, and Calabria, Stanford and George Mason Universities, and the Swiss Federal Institute of Technology in Zurich.